Highlights

**Tropical Support Vector Machines: Evaluations and Extension to Function Spaces**

Ruriko Yoshida, Misaki Takamori, Hideyuki Matsumoto, Keiji Miura

- We obtained generalization error bounds of tropical Support Vector Machines (SVMs) via the Vapnik-Chervonenkis dimensions of tropical hyperplanes using Tropical Radon Lemma by Jaggi et al. (2008).

- We demonstrated theoretically by extreme value statistics that the tropical SVMs are robust against the curse of dimensionality for classifying data points from two Gaussian distributions as well as experimentally recorded activities from different neuron types.

- We defined tropical SVMs over a function space to enable the classification of curves, which is a task we encounter frequently in practice, such as classifications on neuronal tuning curves.
Tropical Support Vector Machines: Evaluations and Extension to Function Spaces

Ruriko Yoshida\textsuperscript{a}, Misaki Takamori\textsuperscript{b}, Hideyuki Matsumoto\textsuperscript{c} and Keiji Miura\textsuperscript{b,∗}

\textsuperscript{a}Department of Operations Research, Naval Postgraduate School, Monterey, 93943, CA, USA
\textsuperscript{b}Graduate School of Science and Technology, Kwansei Gakuin University, Sanda, 669-1337, Hyogo, Japan
\textsuperscript{c}Graduate School of Medicine, Osaka City University, 558-8585, Osaka, Japan

\textbf{ARTICLE INFO}

\textbf{Keywords:}
Extreme Value Statistics
Function Spaces
Max-plus Algebra
Supervised Learning
Tropical Geometry

\textbf{ABSTRACT}

Support Vector Machines (SVMs) are one of the most popular supervised learning models to classify using a hyperplane in an Euclidean space. Similar to SVMs, tropical SVMs classify data points using a tropical hyperplane under the tropical metric with the max-plus algebra. In this paper, first we show generalization error bounds of tropical SVMs over the tropical projective torus. While the generalization error bounds attained via Vapnik-Chervonenkis (VC) dimensions in a distribution-free manner still depend on the dimension, we also show numerically and theoretically by extreme value statistics that the tropical SVMs for classifying data points from two Gaussian distributions as well as empirical data sets of different neuron types are fairly robust against the curse of dimensionality. Extreme value statistics also underlie the anomalous scaling behaviors of the tropical distance between random vectors with additional noise dimensions. Finally, we define tropical SVMs over a function space with the tropical metric.

1. Introduction

In data science, one of the well-known challenges we face is to classify data points with a large number of predictors. For example, in image processing, in order to discriminate one object, such as missiles or face recognition, from others in images, these images are described as a vector in pixels which are typically in a very high dimensional vector space (Hung-Chih Chiang, Moses and Potter (2000)). In bioinformatics, researchers try to classify particular diseases using high dimensional data sets such as micorarrays or SNPs (Fan and Ren (2006)). Fan and Ren (2006) summarize challenges and difficulties with high dimensionality in classification.

Support Vector Machines (SVMs) are one of well-known supervised learning models to classify data points using a hyperplane and introduced by Boser, Guyon and Vapnik (1992) and Cortes and Vapnik (1995). The classical SVMs introduced by Boser et al. (1992) can be written as the \( L_2 \) norm SVMs, that is a hinge loss plus the \( L_2 \) norm penalty formulation. Zhang and Zhou (2010) showed that including many redundant features can cause difficulties in the performance of \( L_2 \) norm SVM. Bradley and Mangasarian (1998) showed that the SVM with the \( L_1 \) norm penalty instead of the \( L_2 \) norm penalty works well with variable selection as well as classification at the same time. This is called \( L_1 \) norm SVMs. While there are several frameworks to analyze the generalization performance, such as covering numbers (Mohri, Rostamizadeh and Talwalkar, 2018), real log canonical thresholds (Hayashi and Watanabe, 2017), mean-field regime (Nitanda, Wu and Suzuki, 2021), and Langevin dynamics regime (Suzuki, 2020), the SVMs especially allow us to derive the upper bounds for the generalization error via Vapnik-Chervonenkis (VC) dimensions (Vapnik, 2000). Although the bounds via VC dimensions may not necessarily be tight, it is still significant to have the first bound for a new variant of SVMs. Peng, Wang and Wu (2016) showed the generalization bound on the probability of errors of the \( L_1 \) norm SVMs and it still depends on the number of predictors, i.e., the dimension of the normal vector of the hyperplane.

Because of advances in computational algebraic geometry (Ren (2015)), tropical geometry finds applications in data science. For example, it can be applied to principal component analysis (Page, Yoshida and Zhang (2020); Yoshida, Zhang and Zhang (2019)) and Bayesian Networks (Améndola, Klüppelberg, Lauritzen and Tran (2020)). With these statistical methodologies, tropical geometry is applied to phylogenomics analyses on Apicomplexa, African coelacanth whole genome data sets, and 1089 full length sequences of hemagglutinin (HA) for influenza A H3N2 from 1993 to 2017 in the state of New York obtained from the GI-SAID EpiFlu (Page et al. (2020); Yoshida et al. (2019)). Gärtner and Jaggi (2006) applied tropical geometry to SVMs. Instead of using a hyperplane defined by the \( L_2 \) metric, Gärtner and Jaggi (2006) used a tropical hyperplane with the tropical metric with the min-plus algebra to classify the data points. Same as the \( L_2 \) norm SVMs, tropical SVMs are the tropicalization of the \( L_2 \) norm hard margin SVMs which maximizes the margin, the distance from the tropical hyperplane to the closest data points in terms of the min-plus algebra. Gärtner and Jaggi (2006) also showed that a hard margin tropical SVM can be formulated as a linear programming problem. Then, Tang, Wang and Yoshida (2020)
showed necessary and sufficient conditions of the optimal solutions of the linear programming problem for finding a hard margin tropical SVM if it is feasible in terms of the max-plus algebra. They also introduced soft margin tropical SVMs and showed that finding a soft margin tropical SVM can be formulated as a linear programming problem.

While there are some developments in computational sides of tropical SVMs, there has not been much in its theoretical evaluation and statistical analysis. For example, it is unclear how the tropical metric with the max-plus algebra. Like the tuning curves (Dayan and Abbott (2001)), instead of vectors.

In this paper, we focus on hard margin tropical SVMs with the max-plus algebra. Like the $L_2$ norm and the $L_1$ norm SVMs, we assume that there exists the optimal tropical hyperplane defined by the normal vector $a \in \mathbb{R}^d/\mathbb{R}I$ such that the probability of the loss function being positive equals to zero. In addition, in order for the tropical metric to be well defined, we consider the tropical projective torus, that is, the projective torus $\mathbb{R}^d/\mathbb{R}I$, where $a := (1, 1, \ldots, 1)$ is defined as the all-one vector.

Note that over the tropical projective torus $\mathbb{R}^d/\mathbb{R}I$ is equal to $(v_1 + c, \ldots, v_d + c)$ with any scalar $c \in \mathbb{R}$ and $\mathbb{R}^d/\mathbb{R}I$ is isometric to $\mathbb{R}^{d-1}$. In applications, the tropical projective torus is useful for subtracting the baseline $\mathbb{R}$ components from feature vectors, as we will see.

Our main contribution of this paper primarily consists of two items: (1) evaluation of generalization errors for tropical SVMs using VC dimensions for tropical hyperplanes over $\mathbb{R}^d/\mathbb{R}I$; and (2) extension to tropical SVMs on a function space, which consists of the tropical metric and a set of functions from a multi-dimensional vector space to a real number.

While generalization error bounds of tropical SVMs over $\mathbb{R}^d/\mathbb{R}I$ still depends on the dimension, our simulations with a set of Gaussian distribution functions show that errors rates of tropical SVMs on $\mathbb{R}^d/\mathbb{R}I$ grow much slower than ones with $L_2$ norm SVMs when we increase the number of predictors while we fix the sample sizes. Then we extend a notion of tropical SVMs to a function space with the tropical metric. In fact, we show that a set of all functions from a multi-dimensional vector space to a real number with the tropical metric is a normed vector space.

This manuscript is organized as follows: In Section 2, we set up basics in tropical geometry under the max-plus algebra. In Section 3, as an application, we show anomalous scaling behaviors of tropical distances. In Section 4, we set up tropical SVMs with the tropical metric over the tropical projective torus as linear programming problems and show the generalization error bounds via the VC dimension of a tropical hyperplane. In Section 5, as an application, we show theoretically by the extreme value statistics the robustness of the tropical SVMs against the curse of dimensionality in a case of two different multivariate Gaussian distributions and empirical data set of neuron types. In Section 6, we present the tropical distance over a function space. Then we show that the tropical distance between functions is metric and we also show that a function space with the tropical metric forms a normed vector space. In addition, we define tropical SVMs over a function space with the tropical metric.

2. Definitions of Tropical Distances and Hyperplanes

In this section, we remind readers some basics in tropical arithmetic and algebra using the max-plus algebra. Through this paper, we consider the tropical projective torus $\mathbb{R}^d/\mathbb{R}I$ which is isometric to $\mathbb{R}^{d-1}$. For more details, see Joswig (2022); Maclagan and Sturmfels (2015); Tang et al. (2020).

Definition 2.1 (Tropical Arithmetic Operations). In this tropical semiring $(\mathbb{R} \cup \{-\infty\}, \oplus, \odot)$, the basic tropical arithmetic operations of addition and multiplication are defined as:

\[
a \boxplus b := \max \{a, b\}, \quad a \odot b := a + b \quad \text{where} \quad a, b \in \mathbb{R} \cup \{-\infty\}.
\]

Over this semiring, $-\infty$ is the identity element under addition and $0$ is the identity element under multiplication.

Definition 2.2 (Tropical Scalar Multiplication and Vector Addition). For any scalars $a, b \in \mathbb{R} \cup \{-\infty\}$ and for any vectors $v = (v_1, \ldots, v_d), w = (w_1, \ldots, w_d) \in (\mathbb{R} \cup \{-\infty\})^d$, we define tropical scalar multiplication and tropical vector addition as follows:

\[
a \odot v \boxplus b \odot w := (\max \{a+v_1, b+w_1\}, \ldots, \max \{a+v_d, b+w_d\}).
\]

Definition 2.3. Suppose $S \subset \mathbb{R}^d/\mathbb{R}I$. If

\[
a \odot v \boxplus b \odot w \in S
\]

for any $a, b \in \mathbb{R}$ and for any $v, w \in S$, then $S$ is called tropically convex.

Definition 2.4 (Tropical Convex Hull). Suppose we have a finite subset $V = \{v^1, \ldots, v^t\} \subset \mathbb{R}^d/\mathbb{R}I$. The tropical convex hull or tropical polytope of $V$ is the smallest tropically-convex subset containing $V$. It can be written as the set of all tropical linear combinations of $V$ such that:

\[\text{tconv}(V) = \{a_1 \odot v^1 \boxplus a_2 \odot v^2 \boxplus \cdots \boxplus a_t \odot v^t \mid a_1, \ldots, a_t \in \mathbb{R}\}.
\]

A tropical polytope of a set of two points $\{v^1, v^2\} \subset \mathbb{R}^d/\mathbb{R}I$ is called a tropical line segment between two points $v^1, v^2$.

Note that over the tropical projective torus $\mathbb{R}^d/\mathbb{R}I$, any point $x = (x_1, \ldots, x_d) \in \mathbb{R}^d/\mathbb{R}I$ can be also written as

\[
x = (x_1 + c, \ldots, x_d + c)
\]

for any $c \in \mathbb{R}$ by the definition of taking mod by $1 := (1, \ldots, 1)$. Therefore, we can assume that

\[
x = (x_1 - x_d, \ldots, x_{d-1} - x_d, 0)
\]
for any point \( x := (x_1, \ldots, x_d) \in \mathbb{R}^d/\mathbb{R}^1 \). Therefore over this manuscript we assume that the last coordinate of any point in the tropical projective torus \( \mathbb{R}^d/\mathbb{R}^1 \) is equal to 0 by taking this normalization.

**Example 2.5** (Examples of Tropical Line Segments). Suppose we have \((2, 3, 0), (2, 1, 0), (2, -1, 0) \in \mathbb{R}^d/\mathbb{R}^1 \). Then, the tropical line segments between the origin \((0, 0, 0) \in \mathbb{R}^d/\mathbb{R}^1 \) and these three points in \( \mathbb{R}^d/\mathbb{R}^1 \) are drawn by blue, green, and red lines in Figure 1.

**Definition 2.6** (Generalized Hilbert Projective Metric). For any \( v, w \in \mathbb{R}^d/\mathbb{R}^1 \) such that \( v = (v_1, \ldots, v_d) \) and \( w = (w_1, \ldots, w_d) \), the tropical distance \( d_{tr} \) between them is defined such that:
\[
d_{tr}(v, w) := \max_i \{v_i - w_i\} - \min_i \{v_i - w_i\}.
\]

**Remark 2.7.** The tropical metric \( d_{tr} \) is a metric over \( \mathbb{R}^d/\mathbb{R}^1 \).

**Example 2.8.** The tropical distances from the origin \((0, 0, 0) \) to the three example points in Figure 1 in \( \mathbb{R}^3/\mathbb{R}^1 \) are
\[
d_{tr}((2, 3, 0), (0, 0, 0)) = 3 - 0 = 3,
\]
\[
d_{tr}((2, 1, 0), (0, 0, 0)) = 2 - 0 = 2,
\]
\[
d_{tr}((2, -1, 0), (0, 0, 0)) = 2 - (-1) = 3.
\]

As one of many geodesics (shortest paths) toward the blue (green, red) point, the tropical line segment is denoted by the blue (green, red) line. Intuitively, a tropical distance is a shortest path length as far as you are allowed to go only parallel to the dotted lines.

**Definition 2.9** (Tropical Hyperplane, Joswig (2005)). For any \( \omega := (\omega_1, \ldots, \omega_d) \in \mathbb{R}^d/\mathbb{R}^1 \), the tropical hyperplane defined by \( \omega \) is the set of points \( x \in \mathbb{R}^d/\mathbb{R}^1 \) such that
\[
H_{\omega} := \left\{ (x_1, \ldots, x_d) \in \mathbb{R}^d/\mathbb{R}^1 | \exists i, j \in \{1, \ldots, d\} \text{ such that } \omega_i + x_i = \omega_j + x_j \geq \max\{\omega_1 + x_1, \ldots, \omega_d + x_d\} \right\}.
\]

In addition, we call \( \omega \) the normal vector of the tropical hyperplane \( H_{\omega} \).

**Remark 2.10.** In terms of tropical geometry, \( H_{\omega} \) is the solutions of the tropical linear function with unknown \( x_1, \ldots, x_d \) such that for a fixed \( \omega = (\omega_1, \ldots, \omega_d) \in \mathbb{R}^d/\mathbb{R}^1 \).
\[
\omega_1 \odot x_1 \odot \ldots \odot \omega_d \odot x_d \quad \text{for} \quad x = (x_1, \ldots, x_d) \in \mathbb{R}^d/\mathbb{R}^1,
\]
which is a special case of tropical polynomials. For more details on finding solutions of a tropical polynomial, see Maclagan and Sturmfels (2015).

**Example 2.11** (Tropical Hyperplane in \( \mathbb{R}^3/\mathbb{R}^1 \)). Suppose we have \( \omega = (1, 2, 0) \in \mathbb{R}^3/\mathbb{R}^1 \). Then
\[
H_{\omega} := \left\{ (x_1, x_2, x_3) \in \mathbb{R}^3/\mathbb{R}^1 | 1 + x_1 = 2 + x_2 \geq x_3 \text{ or } 1 + x_1 = x_3 \geq 2 + x_2 \text{ or } 2 + x_2 = x_3 \geq 1 + x_1 \right\}.
\]

This means that we have four cases:

1. **Case 1:** The first term and the second term are equal and max. \( 1 + x_1 = 2 + x_2 \geq x_3 = 0 \).
2. **Case 2:** The first term and the third term are equal and max. \( 2 + x_2 < 1 + x_1 = x_3 = 0 \).
3. **Case 3:** The second term and the third term are equal and max. \( 1 + x_1 < 2 + x_2 = x_3 = 0 \).
4. **Case 4:** All terms are equal and max. \( 1 + x_1 = 2 + x_2 = x_3 = 0 \).

We can set \( x_3 = 0 \) without loss of generality, since \((x_1, x_2, x_3) \) and \((x_1 - x_3, x_2 - x_3, 0) \) represent the same point in \( \mathbb{R}^3/\mathbb{R}^1 \). Thus \( H_{(1,2,0)} \) is the set of \((x_1, x_2, 0) \) on the three half lines in Figure 2.

**Definition 2.12** (Sectors of Tropical Hyperplane, Joswig (2005)). Each tropical hyperplane \( H_{\omega} \) divides the tropical projective torus \( \mathbb{R}^d/\mathbb{R}^1 \) into \( n \) connected components. These connected components are open sectors defined as:
\[
S^i_{\omega} := \{ x \in \mathbb{R}^d/\mathbb{R}^1 | \omega_i + x_i > \omega_j + x_j, \forall j \neq i \}, \quad i = 1, \ldots, d.
\]

**Definition 2.13** (Tropical Distance to a Tropical Hyperplane). The tropical distance \( d_{tr} \) from a point \( x \in \mathbb{R}^d/\mathbb{R}^1 \) to a tropical hyperplane \( H_{\omega} \) is:
\[
d_{tr}(x, H_{\omega}) := \min \{d_{tr}(x, y) | y \in H_{\omega} \}.
\]
Proposition 2.14 (Lemma 2.1 in Gärtner and Jaggi (2006)). Let $H_0$ be the tropical hyperplane defined by the zero vector $0 = (0, 0, \ldots, 0) \in \mathbb{R}^d / \mathbb{R}$. For any $x = (x_1, \ldots, x_d) \in \mathbb{R}^d / \mathbb{R}$,

$$d_{\text{tr}}(x, H_0) = \max(x) - \text{second max}(x).$$

Corollary 2.15 (Corollary 2.3 in Gärtner and Jaggi (2006)). For any $\omega \in \mathbb{R}^d / \mathbb{R}$ and for any $x \in \mathbb{R}^d / \mathbb{R}$,

$$d_{\text{tr}}(x, H_0) = d_{\text{tr}}(\omega + x, H_0).$$

Example 2.16. We use the sample $H_0$ from Example 2.11. Let $d = 3$ and suppose $\omega = (1, 2, 0) \in \mathbb{R}^3 / \mathbb{R}$. In addition, suppose we have a point $x = (1, 1, 0) \in \mathbb{R}^3 / \mathbb{R}$. By Corollary 2.15,

$$d_{\text{tr}}(x, H_0) = d_{\text{tr}}(x + \omega, H_0) = d_{\text{tr}}((2, 3, 0), H_0) = 3 - 2 = 1.$$

3. Application 1: Valuable Scaling Behaviors of Tropical Distances

3.1. Bell-shaped Tuning Curves

To characterize the properties of the tropical distances $d_{\text{tr}}(v, w)$ proposed in this paper, we computed the tropical distances for the mathematical models of neuronal tuning curves as benchmarks. The tuning curves are the vectors consisting of neural responses to different stimuli. For example, the responses of neurons in the primary visual cortex to oriented bars on the screens as visual stimuli are known to show famous bell-shaped curves as a function of the orientation.

We modeled neuronal tuning curves basically as standard Gaussian distribution functions and examined how the distances scale when one of two curves is horizontally shifted or vertically lifted. Both the tropical and Euclidean distances between two bell-shaped tuning curves increased with the increasing horizontal shift between them (Figure 3, left). Meanwhile, the tropical distances between the two bell-shaped tuning curves was insensitive to the vertical lift of tuning curves, while the Euclidean distance was not (Figure 3, right). The insensitivity to the vertical shift originates from the definition of the tropical distances $d_{\text{tr}}(v, w)$ where tuning curves are considered in $\mathbb{R}^d / \mathbb{R}$ and $(1, 1, \ldots, 1)$-direction is collapsed and ignored in Def. 2.6. That is, $d_{\text{tr}}(v + c1, w + d1) = d_{\text{tr}}(v, w)$ for any $c$ and $d$ in $\mathbb{R}$. Note that the entire lifting of tuning curves are often observed as a result of drifting background neural activities (Okun, Steinmetz, Cossef, Iacaruso, Ko, Barthó, Moore, Hofer, Mrsic-Flogel, Carandini and Harris (2015); Stringer, Pachitariu, Steinmetz, Reddy, Carandini and Harris (2019); Miura, Tsubo, Okada and Fukai (2007); Miura (2013); Takahashi, Maruyama, Ito and Miura (2019)). Thus, practically, the tropical distance can be useful for the sake of ignoring the temporal drift of tuning curves.

3.2. Random Tuning Curves

Next we considered a random tuning curve consisting of noisy responses. When we computed the distances between a random tuning curve

$$v = (5, -x_2, -x_3, \ldots, -x_n),$$

where $x_i \sim \text{Exp}(1)$ and a flat tuning curve,

$$w = (0, 0, 0, \ldots, 0),$$

the tropical distance was relatively insensitive to the additional noise dimensions. That is, the tropical distance scaled with $\log n$ while the Euclidean distance scaled with $\sqrt{n}$. The anomalous scaling for the tropical distance can be explained by the extreme value statistics (Gumbel (2004)) as

$$d_{\text{tr}}(v, w) := \max_i (v_i - w_i) - \min_i (v_i - w_i) \approx 5 + \max_i x_i.$$
Then, because \( \max \{ x_i \} - \log n \sim \text{Gumbel}(0,1) \) (Gumbel (2004)), the expectation is given as

\[
d_{\text{tr}}(v, w) = 5 + \gamma + \log n. \tag{4}
\]

This theoretical result explains the simulation result in Figure 4. In this way, the extreme value statistics plays the crucial role in the computation of the tropical distance between random vectors.

4. Evaluation of Tropical Support Vector Machines (SVMs)

4.1. \( L_2 \)-norm SVMs

In this section we provide an overview of the \( L_2 \) norm SVMs. Let \( D_2 \) be the distribution on the joint random variable \((X, Y)\) for \( X \in \mathbb{R}^d \) and \( Y \in \{-1,1\} \) and let \( S_2 \) be the sample \( S_2 := \{(X_1, Y_1), \ldots, (X^n, Y^n)\} \). Then the standard linear SVM is a statistical model by solving the following regularization problem:

\[
\min_{\omega} \left( \lambda ||\omega||^2 + n^{-1} \sum_{i=1}^{n} (1 - Y^i ((X^i)^T \omega + o_0))^+ \right),
\]

where \( ||x|| \) is the \( L_2 \) norm of a vector \( x \in \mathbb{R}^d \), \( (1 - u)_+ = \max\{1 - u, 0\} \) is the hinge loss function, \( \lambda \) is a tuning parameter, \( o_0 \) is a constant term of the hyperplane, and \( \omega \) is the normal vector of the hyperplane to separate the data points. Vapnik (2000) discussed the generalization bounds on the \( L_2 \) norm SVMs using Theorem 4.11 and the VC dimension of the \( L_2 \) norm SVMs. In addition, using the Rademacher complexity of \( L_2 \) norm of unit vectors (Bartlett and Mendelson (2003)), one can show the following generalization bound of the error rate:

**Theorem 4.1.** Let \( \omega^* \) be an output from the classical hard margin SVM from a sample \( S \) with the distribution \( \pi_+ = P(Y = 1) \) and \( \pi_- = P(Y = -1) \). Let \( X \in \mathbb{R}^d \) be a random variable and \( D_2 \) be the distribution for a random variable \((X, Y)\). Let \( ||x|| \) be the \( L_2 \) norm of a vector \( x \in \mathbb{R}^d \). Assuming that \( ||X|| \leq R \) with probability 1 and there exists \( \omega^* \) with \( P_{D_2}(Y^i((X^i)^T \omega^* + o_0)) \geq 1 \) = 1. Then, for any \( \eta > 0 \) with the probability greater than or equal to \( 1 - \eta \) we have

\[
P_{D_2}(Y \neq \text{sign}(X \cdot \omega^*)) \leq \frac{4R||\omega^*||}{\sqrt{n}} + (1 + 2R||\omega^*||)\sqrt{\frac{2\log(4||\omega^*||/\eta)}{n}}. \tag{5}
\]

There has been much work done in tighter generalization bounds on the \( L_2 \) norm SVMs using the VC dimension. As it is not this paper’s focus to discuss the details on the generalization bounds for the \( L_2 \) norm SVMs, for more details, see Burges (1998); Guermeur (2007) and the references therein.

**Remark 4.2.** Recall that the \( L_p \) norm SVMs, for \( 0 \leq p \leq \infty \), can be written as the following optimization problem:

\[
\min_{\omega} \left( \lambda ||\omega||^p + n^{-1} \sum_{i=1}^{n} (1 - Y^i ((X^i)^T \omega + o_0))^+ \right),
\]

where \( ||x||_p \) is the \( L_p \) norm of a vector \( x \in \mathbb{R}^d \), \( (1 - u)_+ = \max\{1 - u, 0\} \) is the hinge loss function, \( \lambda \) is a tuning parameter, and \( \omega \) is the norm vector of the hyperplane to separate the data points.

4.2. Tropical SVMs

In this section we provide an overview the hard margin tropical SVMs originally developed by Gärtnert and Jaggi (2006) with the random variable \( X \in \mathbb{R}^d/\mathbb{R} \) given the response variable \( Y \in \{0,1\}^d \) for one-hot encoding. Note that tropical SVMs can perform multiclass classification naturally. Before we formally define the hard margin tropical SVMs, we need to define some notation.

Let \( S(x) \) be a set of indices of nonzero elements in a vector \( x \in \mathbb{R}^d/\mathbb{R} \), i.e., \( S(x) \subset \{1, \ldots, d\} \) where \( x_i \neq 0 \). Let \( I_{\omega}(x) \in \{0,1\}^d \) be a vector of indicator functions of index set \( \{1, \ldots, d\} \) of a vector \( x \) with a vector \( \omega \in \mathbb{R}^d/\mathbb{R} \) where

\[
I_{\omega}(x)_i = \begin{cases} 1 & \text{if } x_i + \omega_i = \max(x + \omega) \\ 0 & \text{otherwise.} \end{cases}
\]

Let \( J_{\omega}(x) \in \{0,1\}^d \) be also a vector of indicator functions of index set \( \{1, \ldots, d\} \) of a vector \( x \) such that

\[
J_{\omega}(x)_i = \begin{cases} 1 & \text{if } x_i + \omega_i = \text{second max}(x + \omega) \\ 0 & \text{otherwise.} \end{cases}
\]

Suppose we have a categorical response variable \( Y' \in \{g_1, \ldots, g_q\} \) with \( q \leq d \) many levels. Then we set a vector of indicator functions \( Y = (Y_1, \ldots, Y_d) \) such that

\[
Y_i = \begin{cases} 1 & \text{if } Y_i = g_i \\ 0 & \text{else.} \end{cases}
\]

**Example 4.3.** Suppose that we have a binary categorical response variable \( Y' \in \{"yes", "no"\} \). Then we can set as a vector of indicator functions \( Y = (Y_1, \ldots, Y_d) \) such that

\[
Y_1 = \begin{cases} 1 & \text{if } Y' = \text{yes} \\ 0 & \text{else,} \end{cases}
\]

\[
Y_2 = \begin{cases} 1 & \text{if } Y' = \text{no} \\ 0 & \text{else,} \end{cases}
\]

and

\[
Y_3 = Y_4 = \ldots = Y_d = 0.
\]

Yoshida et al.: Preprint submitted to Elsevier
In order to simplify the problem, here we consider a case that there are only two classes $Y^A$ and $Y^B$ in the response variable. More precisely, we have a random variable $Y^A := (Y^A_1, \ldots, Y^A_d), Y^B := (Y^B_1, \ldots, Y^B_d) \in \{0, 1\}^d$ such that for fixed $i, k \in \{1, \ldots, d\}$ with $i \neq k$, $Y^A_j = \begin{cases} 1 & \text{if } j = i \\ 0 & \text{else} \end{cases}$ and $Y^B_l = \begin{cases} 1 & \text{if } l = k \\ 0 & \text{else} \end{cases}$ with a discrete probability $\pi_A := P(Y = Y^A)$ and $\pi_B := P(Y = Y^B)$. Then, suppose we have a multivariate random variable $X \in \mathbb{R}^d/\mathbb{R}I$ given $Y$ with the probability density function $f_A$ if $Y = Y^A$ and the probability density function $f_B$ if $Y = Y^B$ such that there exists a tropical hyperplane $H_\omega^*$ with a normal vector $\omega^* \in \mathbb{R}^d/\mathbb{R}I$ with the following properties:

(i) there exists an index $i \in \{1, \ldots, d\}$ such that for any $j \in \{1, \ldots, d\} \setminus \{i\}$, $\omega^*_i + X_i > \omega^*_j + X_j$, and

(ii) $\max \{I_\omega^*(X) - Y\} = 0$,

with probability 1.

Let $D$ be the distribution on the joint random variable $(X, Y)$ and let $S$ be the sample $S := \{(X^1, Y^1), \ldots, (X^n, Y^n)\}$. Then we formulate an optimization problem for solving the normal vector $\omega$ of an optimal tropical separating hyperplane $H_\omega$ for random variables $X$ given $Y$: For some cost $C \in \mathbb{R}$ $\max_{c \in \mathbb{R}^d/\mathbb{R}I} \sum_{i=1}^{\max\{I_\omega(X) - Y\}} \left(\max_{\omega_i} \{I_\omega(X) - Y\} + \frac{C}{\min\{\omega_i\}}\right)$

Here, the expectation of the random variable $\max \{I_\omega(X) - Y\}$ is the 0-1 loss function. Also note that $d_i(X, H_\omega) = X_i + \omega_i - X_j - \omega_j$, where $i \in S(I_\omega(X), j \in S(I_j(X))$. Thus, this optimization problem can be explicitly written as a linear programming problem (7)-(10) below, where the optimal solution $z$ means the margin of the tropical SVM: For some cost $C \in \mathbb{R}$

\begin{align*}
\max_{z, c} \sum_{i=1}^{\max\{I_\omega(X) - Y\}} \left(z + C \frac{\min\{I_\omega(X)\}}{\omega_i}\right) \\
\text{s.t. } \forall X \in S, \forall i \in S(I_\omega(X), z + X_i + \omega_i - X_j - \omega_j \leq 0, \\
\forall X \in S, \forall i \in S(I_\omega(X), \omega_i - \omega_j \leq X_i - X_j, \\
\forall X \in S, \forall i \in S(I_\omega(X), \omega_i - \omega_j \leq X_i - X_j, \\
\forall X \in S, \forall i \in S(I_\omega(X), \omega_i - \omega_j \leq X_i - X_j.
\end{align*}

Remark 4.4. With the regularization with the tropical metric to the tropical hyperplane, it minimizes the coefficients of the tropical hyperplane $\omega_i$ for $i = 1, \ldots, d$ and tries to minimize the difference between the coordinate of the maximum and second maximum of $X + \omega$. This means that this regularization behaves similar to the $L_1$ or $L_\infty$ norm regularization, that is, to select features which contribute to discriminate $X|Y^A$ and $X|Y^B$.

4.3. Bound on Generalization Errors via VC Dimensions

In this section we show the generalization bound for the error rate of the hard margin tropical SVM. In order to compute them, we define the loss function in terms of the distribution of $X|Y$ defined in the previous section and the sample loss function. In order to prove the generalization bound, we use the VC dimension (Vapnik, 1995; Mohri et al., 2018)

Let $D$ be the distribution of $X|Y$ defined in Section 4. In addition, let

\begin{align*}
L_D(\omega) &= \mathbb{E}_D \left( \max \{I_\omega(X) - Y\} \right) \\
L_S(\omega) &= \mathbb{E}_S \left( \max \{I_\omega(X) - Y\} \right)
\end{align*}

be the loss function with respect to $D$ and let $L_S(\omega) = \mathbb{E}_S \left( \max \{I_\omega(X) - Y\} \right)$ be the loss function with respect to the sample $S$.

Let $(Z, \omega^*)$ be an output of the tropical SVM. Then we want to find the upper bound for

\begin{align*}
P \left( \max \{I_\omega(X) - Y\} > 0 \right) &= L_D(\omega^*).
\end{align*}

In order to prove our bound, we have to define VC dimension.

Definition 4.5. Suppose $M$ is a set and a family of a subset $R \subset 2^M$. A set $T \subset M$ is called shattered by $R$ if $|\{r \cap T | r \in R\}| = 2^{|T|}$.

The pair $(M, R)$ is called range space. The VC dimension of a range space, $VC - \dim(M, R)$ is the maximal cardinality of a subset of $M$ which can be shattered by $R$.

Example 4.6 (Linear Classifiers). In $(d - 1)$-dimensional Euclidean space, a linear classifier can shatter or successfully classify $d$ points in general positions with arbitrary labels. Thus, the VC dimension of the linear classifier is $d$.

Lemma 4.7 (Tropical Radon Lemma in Jaggi, Katz and Wagner (2008)). For $d + 1$ points in $\mathbb{R}^d/\mathbb{R}1$, there exists a partition into two sets with intersecting convex hulls, which are therefore not separable by any tropical hyperplane $H_\omega$.

Example 4.8. Any tropical hyperplane in $\mathbb{R}^3/\mathbb{R}1$ cannot shatter the four points shown in Figure 5. The fact that the two line segments intersect to each other suggests that it is impossible to classify them by tropical hyperplanes. In fact, ANY four points cannot be shattered by a tropical hyperplane in $\mathbb{R}^2/\mathbb{R}1$. This is because, given four or more points, there always exist ways of partitioning so that the tropical convex hull of the points in each class intersects.

Lemma 4.9 (Lemma 27 in Jaggi et al. (2008)). There exists a set of $d$ points in $\mathbb{R}^d/\mathbb{R}1$ that can be shattered by single sectors defined by a tropical hyperplane $H_\omega$. 

Yoshida et al.: Preprint submitted to Elsevier

Page 6 of 13
While Tropical Radon Lemma 4.7 shows that the VC dimension of a tropical hyperplane is smaller than \( d + 1 \), (trivial) Lemma 4.9 shows that it is larger than or equal to \( d \) (as single sectors should have less ability than hyperplanes).

**Theorem 4.10** (Jaggi et al. (2008)). *Let \( \mathcal{H} \) be a family of tropical hyperplanes in \( \mathbb{R}^d \). Then the VC dimension of a range space \((\mathbb{R}^d / \mathbb{R}1, \mathcal{H})\) is \( d \).*

**Theorem 4.11** (Vapnik (2000)). *Let \( \mathcal{H} \) be a family of classification models whose VC dimension is \( h \). Let \( n \) be the sample size of the sample \( S \). Then, for any \( \eta > 0 \) and for a sufficiently large sample size \( n \) with the probability greater than or equal to \( 1 - \eta \), we have

\[
\ell_D(H) - \ell_S(H) \leq \frac{n}{h} (\log(2n/h) + 1) - \log(\eta/4),
\]

where \( \ell_D(H) \) is the \( 0 \)-\( 1 \) loss for the distribution \( D \) with \( \mathcal{H} \) and \( \ell_S(H) \) is the \( 0 \)-\( 1 \) loss for the sample \( S \) with \( \mathcal{H} \).

The generalization error bound for tropical SVMs is obtained by plugging in the VC dimension of tropical hyperplanes in Theorem 4.10 for \( h \) in Theorem 4.11.

**Theorem 4.12.** *Let \( n \) be the sample size of the sample \( S \). Then, for any \( \eta > 0 \) and for a sufficiently large sample size \( n \) with the probability greater than or equal to \( 1 - \eta \), we have

\[
P_D \left( \max \left\{ I_\omega(X) - Y \right\} > 0 \right) \leq \sqrt{\frac{d (\log(2n/d) + 1) - \log(\eta/4)}{n}}.
\]

**Proof.** Let \( S \) be a sample for a test set. Since there is a tropical hard margin and \((z_s, \omega^s)\) is a feasible solution on the test set \( S \), \( L_S(\omega^s) = 0 \). With Theorem 4.10 and Theorem 4.11, we have

\[
L_D(\omega^s) - L_S(\omega^s) \leq \sqrt{\frac{d (\log(2n/d) + 1) - \log(\eta/4)}{n}}.
\]

Since \( L_S(\omega^s) = 0 \) by the assumption and

\[
L_D(\omega^s) = P_D \left( \max \left\{ I_\omega(X) - Y \right\} > 0 \right),
\]

we have the result. \( \square \)

5. Application 2: Robustness of Tropical SVMs against Curse of Dimensionality

In the previous section, we derive the generalization error bound for tropical SVMs. The bound is stated in a distribution-free manner as a general feature applicable to all data. Specifically, the bound is derived purely combinatorially only based on the shapes of the hyperplanes. Algorithmic details of tropical SVMs are not yet taken into account there. Therefore it is possible that tropical SVMs outperform in specific situations, for example, when only a few features are prominently informative among many. Thus the evaluation of the performance of tropical SVMs by computational experiments can give complementary information. Here we applied tropical SVMs to specific problems of computational experiments and real data analyses to evaluate its performance originating from the max-plus algebra.

5.1. Implementation of Tropical SVMs

Soft margin tropical SVMs are defined in Tang et al. (2020) which allow some data points in the wrong open sectors with the hinge loss function as similar to the \( L_2 \) norm SVMs. Since the loss function deals with \( L_2 \) norm of \( I_\omega(X) - Y \), we have to run exponentially many linear programming problems in order to find the optimal tropical hyperplane to fit the data set. Therefore, Tang et al. introduced several heuristic algorithms to estimate the optimal tropical hyperplane to fit the dataset: Algorithms 1 – 4. In the computational experiments in this paper, we applied Algorithms 3, the simplest one, from Tang et al. (2020) to fit each data to the model, as it always performed the best for our simple examples. We obtained the \( r \) code from https://github.com/HoujieWang/Tropical-SVM.

5.2. Computational Experiments

As we observe in Section 3.2, the tropical distance can ignore the nuisance dimensions that are purely noises. Therefore it is expected that tropical SVMs can show a good performance in the curse-of-dimensionality regime where extra dimensions are all uninformative.

Here we performed a simple computational experiment of the two-class classification using the uncorrelated unit-variance Gaussian distributions whose means are \( m_1 = (s, -s, 0, 0, \ldots, 0) \) for class 1 and \( m_2 = (-s, s, 0, 0, \ldots, 0) \) for class 2, where \( s = \sqrt{2} \) or 5. To be specific, the unit covariance matrices are common for the two distributions as \( \Sigma_1 = \Sigma_2 = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \). Note that the signal or the difference of the means, \( m_2 - m_1 = (2s, -2s, 0, 0, \ldots, 0) \), is orthogonal to the irrelevant direction \((1, 1, 1, 1, \ldots, 1)\). As a measure of separation, the d-prime or the Euclidean distance between \( m_1 \) and \( m_2 \) divided by the S.D. (=1) is given by \( d' = 2s\sqrt{2} \).

Before we look at the curse-of-dimensionality regime or small samples, we first consider the case when the numbers of training and test samples \( N \) are fairly large to compare
the numerically obtained generalization errors with the theoretical bounds in Figure 6. We draw the lower bound of the classification success rates by subtracting the penalty term in the right hand side of the inequality in Theorem 4.11 from the hit rate for the training set. The gaps between the numerical computations and the theoretical bounds suggest that the bounds are not so tight. In fact, the lower bounds are mostly smaller than the chance level 50% and thus not informative. Although we solely use $\eta = 0.1$, the bounds are rather insensitive to the value of $\eta$. Even if we have as large as $N = 1000$ or 10000 observations, the gap is still quite noticeable as expected from the penalty term in the right hand side of the inequality in Theorem 4.11.

Next, we consider small samples as a curse-of-dimensionality regime. Figure 7 shows the classification success rates of two multivariate Gaussian distributions separated by $d' = 4$ (top row) or $10\sqrt{2}$ (bottom row) and covariances are the unit matrices. The numbers of training and test samples for each class are 5 (left column) and 10 (right column). The cross-validated performance averaged over 1000 realizations of data were plotted. The errorbars denote the standard deviations for the 1000 realizations. Note that the tropical SVM is robust against the curse of dimensionality.

### 5.3. Understanding Robustness by Numerical Examples

First, as a basic example, let us consider the simplest case with two points in three dimensions: $x_1 = (5, -5, 0)$ for class 1 and $x_2 = (-5, 5, 0)$ for class 2. As our problem is to find $\omega = (\omega_1, \omega_2, \omega_3)$ in $\mathbb{R}^3/\mathbb{R}_1$, we can limit to $\omega = (\omega_1, \omega_2, 0)$ without loss of generality. By maximizing the following margin over $\omega$ in $\mathbb{R}^2/\mathbb{R}_1$ under the condition with no classification error,

$$M(\omega) := \min\left(d_{\text{tr}}(x_1, H_\omega), d_{\text{tr}}(x_2, H_\omega)\right),$$

we obtain $\omega^* = (5 + c, 5 + c, 0)$ for $c > 0$ as in Figure 8.

In fact, $x_1 + \omega^* = (10 + c, c, 0)$ and $x_2 + \omega^* = (c, 10 + c, 0)$ lead to $d_{\text{tr}}(x_1, H_\omega^*) = d_{\text{tr}}(x_2, H_\omega^*) = 10$. In Figure 8 (left) for the margin $M(\omega)$, error-free regions for $\omega$ are only A, B and C. The other regions have classification errors. The entire shape of the error-free region reflect the Y-shape of the hyperplane $H_0$. For example, for $\omega = (-10, -10, 0)$, $x_1 + \omega$ and $x_2 + \omega$ belong to the same bottom left sector of $H_0$, meaning misclassification. Note that as the margin function $M(\omega)$ is convex unless $x_1 + \omega$ or $x_2 + \omega$ crosses the sector borders, the linear programming is needed only once for each possible combination of labels and sectors (Tang et al. 2020).

Here we can impose the regularization term, $d_{\text{tr}}(\omega, 0)$, so that we choose the solution with the minimum norm even if the evaluation function outputs ties. In this tradition, the unique solution is $\omega^* = (5, 5, 0)$. 

---

**Figure 6**: Numerically obtained classification success rates (thick lines) and theoretical lower bounds (dotted lines) for the classification of two multivariate Gaussian distributions whose means are separated by $d' = 4$ and covariances are the unit matrices. The results for tropical SVMs are red colored while those for classical SVMs are black colored. The numbers of training and test samples for each class are 100 (left column) and 200 (right column). The cross-validated performance averaged over 1000 realizations of data were plotted. Note that the lower bounds of the hit rates (=upper bounds of error rates) are not tight.

**Figure 7**: Classification success rates of two multivariate Gaussian distributions whose means are separated by $d' = 4$ (top row) or $10\sqrt{2}$ (bottom row) and covariances are the unit matrices. The numbers of training and test samples for each class are 5 (left column) and 10 (right column). The cross-validated performance averaged over 1000 realizations of data were plotted. The errorbars denote the standard deviations for the 1000 realizations. Note that the tropical SVM is robust against the curse of dimensionality.
Second, let us consider the general case with \( N \) points for each class in \( d \) dimensions, whose coordinates are perturbed from \( x_1 \) or \( x_2 \) by standard Gaussian noises \( \xi \) or \( \eta \) (Figure 9) as in the previous computational experiments. We again need to maximize the margin:

\[
M(\omega) := \min_{1 \leq i \leq N, 1 \leq j \leq 2} d_{ui}(x^i_{y=1}, H_\omega),
\]

with

\[
x^i_{y=1} = (5 + \xi_1^i, -5 + \xi_2^i, \xi_3^i, \xi_4^i, \ldots, \xi_d^i),
\]

\[
x^i_{y=2} = (-5 + \eta_1^i, 5 + \eta_2^i, \eta_3^i, \eta_4^i, \ldots, \eta_d^i),
\]

where \( \xi_p^i \sim N(0, 1) \) and \( \eta_p^i \sim N(0, 1) \) for \( 1 \leq p \leq d \) and \( 1 \leq i \leq 2 \). Let us try to find the solution near \( \omega^* = (5 + c, 5 + c, 0, 0, \ldots, 0) \) and \( d \to \infty \).

In the previous computational experiments, we observed that for all cases of 1000 data realizations the best classifiers used the first and the second sectors as in Figure 9. Furthermore, among \( x_1 + \omega_1 \) for \( 1 \leq p \leq d \), the maximum and the second maximum were attained by \( x_1 + \omega_1 \) and \( x_2 + \omega_2 \) when the training data were classified. Therefore we focus on this situation, i.e., classification by using only sector 1 and sector 2, in what follows.

Under the assumption of \( \omega_2 \gg 5 \), which will be justified later, the tropical distance to \( H_\omega \) is

\[
d_{ui}(x^i_{y=1}, H_\omega) = d_{ui}(x^i_{y=1} + \omega, H_0) = \{ \max - 2\text{nd max of } \} (5 + \xi_1^i + \omega_1, -5 + \xi_2^i + \omega_2, \xi_3^i + \omega_3, \ldots, \xi_d^i + \omega_d)
\]

\[= 10 + \xi_1^i + \omega_1 - \xi_2^i - \omega_2 \quad (16)\]

The minimum over \( N \) points with class 1 is

\[
\min_{1 \leq i \leq N} d_{ui}(x^i_{y=1}, H_\omega) = 10 + \omega_1 - \omega_2 + \min_{1 \leq i \leq N} (\xi_1^i - \xi_2^i)
\]

where \( i_s(\omega) \) denotes the minimizer. Similarly for the class 2,

\[
\min_{1 \leq j \leq N} d_{uj}(x^j_{y=2}, H_\omega) = 10 + \omega_2 - \omega_1 + \min_{1 \leq j \leq N} (\eta_1^j - \eta_2^j).
\]

This theoretical prediction perfectly matched with the computational experiments. This can be rewritten as,

\[
\omega_1 = 5 + c
\]

\[
\omega_2 = 5 + \frac{\xi_1^i - \xi_2^i + \eta_1^j - \eta_2^j}{2},
\]

for \( c > 0 \), which give the (evenly scored) maximizers of the margin under \( \omega_2 \gg 5 \). Note that \( \omega \) is perturbed from \( (5 + c, 5 + c) \) only slightly of order \( O(1) \). This equation illustrates that the hyperplane just listens to the positions of the support vectors. Adding \( c \) might look strange at first glance, but it is interpreted as the hyperplane’s freedom to shift along the irrelevant coordinates of the support vectors. In fact, the hyperplane for classical SVMs actually has the same freedom but you just cannot distinguish the shifted planes.

Now we can explain why tropical SVMs are robust against the curse of dimensionality. In the computational experiments, the other \( \omega_3(p \geq 3) \) were smaller by around five or more. Therefore the misclassification never happened as the maximum of \( (x + \omega) \) always occurred at the first or second coordinate appropriately.
5.4. Real Data Analysis

We now turn to empirical data from brain science. To summarize the data acquisition in this paragraph, we recorded the spiking activity of 493 total neurons in the ventral tegmental area (VTA) and the substantia nigra (SN) of DAT (dopamine transporter)-cre mice using tetrodes while these mice performed various classical conditioning tasks (Matsumoto, Tian, Uchida and Watabe-Uchida (2016)). We optogenetically identified 179 dopaminergic neurons (Cohen, Haesler, Vong, Lowell and Uchida (2012); Matsumoto et al. (2016)), which enabled us to assess the activity before reward stimuli, which are least informative. The following features we preferentially used are the responses to dopaminergic neurons based on their activities as features.

Figure 10: Classification success rates of dopaminergic and nondopaminergic neurons by using various numbers of features for tropical and classical SVMs. All features are the firing rates, i.e. the number of spikes per second, in 300ms bins. The first 7 features we preferentially used are the responses to 7 different reward stimuli, which are informative. The following 7 features we appended are spontaneous activities after reward stimuli, which are less informative. Note that the classification success rates decreased slower for the tropical SVM than the classical SVM, suggesting that the tropical SVM is relatively robust against the curse of dimensionality.

Here we did not necessarily use the entire features but rather changed the numbers of features used for the classification. As the order of adding features can matter, we ordered and added the putatively informative features first. After the classification success rates showed the peaks at the beginning, they decreased slower for the tropical SVM than the classical SVM (Figure 10).

In general, we believe that although the classification success rate is enhanced by adding some most informative features at the beginning, soon it will be deteriorated by less informative features, which can be a majority in real world big data. The tropical SVM may be superior in the latter regime of curse of dimensionality.

6. Extension of Tropical SVMs to Function Spaces

In this section, we define tropical SVMs over a function space to enable the classification of curves, which we encounter frequently in practice. For example, we consider neuronal tuning curves in Figure 3 as a feature vector. That is, we can treat a tuning curve as a point in a function space. For example, it is valuable for identifying dopaminergic neurons in clinical researches (Ishikawa, Matsumoto and Miura, 2018, 2019) to measure the distances between neuronal tuning curves and classify neurons using a tropical hyperplane boundary, which itself can be another curve in tropical SVMs.

6.1. Tropical Distances and Hyperplanes on Function Spaces

Suppose we have a set of functions mapping from \( \mathbb{R}^d \) to \( \mathbb{R} \) denoted as \( F \) such that:

\[
F := \{ f : \mathbb{R}^d \rightarrow \mathbb{R} : |f(x)| \text{ is bounded}, f(x) = f(x) + c, \forall x \in \mathbb{R}^d, \forall c \in \mathbb{R} \}.
\]

Let \( d_{tr} \) be a distance between functions in \( F \) such that:

\[
d_{tr}(f, g) := \max\{ |f(x) - g(x)| : \forall x \in \mathbb{R}^d \} - \min\{ |f(x) - g(x)| : \forall x \in \mathbb{R}^d \}.
\]

Example 6.1 (Functional Tropical Distances, Figure 11).

Let \( F \) be a set of univariate Gaussian distribution functions with \( \mu \) and \( \sigma \), and mixtures of these Gaussian distributions with real valued coefficients. Suppose \( F_1, F_2, F_3 \) and \( F_4 \) are univariate Gaussian distribution functions with \( \mu_1 = \mu_2 = -2, \mu_3 = 2, \sigma_1 = \sigma_3 = 1, \) and \( \sigma_2 = \sigma_4 = 1/2 \). Then, \( d_{tr}(F_1, F_2) = d_{tr}(F_3, F_4) = \frac{11}{8\sqrt{2\pi}} = 0.549 \), \( d_{tr}(F_1, F_3) = 0.798 \), and \( d_{tr}(F_1, F_4) = d_{tr}(F_2, F_3) = 1.197 \).

Lemma 6.2. \( d_{tr}(f, g(x), g(x)) = d_{tr}(-f, -g(x)) = d_{tr}(g(x), f(x)) \)

Proof:

\[
d_{tr}(f, g(x)) = -\min\{ f(x) - g(x) : \forall x \in \mathbb{R}^d \} + \max\{ f(x) - g(x) : \forall x \in \mathbb{R}^d \}
\]

\[
= \max\{ -f(x) + g(x) : \forall x \in \mathbb{R}^d \} - \min\{ -f(x) + g(x) : \forall x \in \mathbb{R}^d \}
\]

\[
= d_{tr}(-f, -g(x))
\]

\[
= d_{tr}(g(x) - f(x), 0)
\]

\[
= d_{tr}(g(x), f(x)).
\]
Lemma 6.3. $d_{it}(f(x), 0)$ is a norm on $F$.

Proof. For the identity,
$$d_{it}(f, 0) = 0 \text{ if and only if } \max\{f(x) : \forall x \in \mathbb{R}^2\} = \min\{f(x) : \forall x \in \mathbb{R}^2\},$$
meaning that $f(x)$ is constant. Next, by Lemma 6.2 with $g(x) = 0$
$$d_{it}(af, 0) = d_{it}(|a|f, 0) = |a|d_{it}(f, 0)$$
for $a \in \mathbb{R}$. For triangle inequality,
$$d_{it}(f(x) + g(x), 0) = \max\{f(x) + g(x) : \forall x \in \mathbb{R}^2\} - \min\{f(x) + g(x) : \forall x \in \mathbb{R}^2\} \leq \max\{f(x) : \forall x \in \mathbb{R}^2\} + \max\{g(x) : \forall x \in \mathbb{R}^2\} = d_{it}(f(x), 0) + d_{it}(g(x), 0).$$

Lemma 6.4. $d_{it}$ is a metric on $F$.

Proof. $d_{it}$ is a metric because it is induced by the norm of the difference,
$$d_{it}(f(x), g(x)) = d_{it}(f(x) - g(x), 0).$$

By the proof of Lemma 6.4, we have the following theorem:

Theorem 6.5. $(F, d_{it})$ is a normed vector space.

Definition 6.6 (Tropical Hyperplane in $F$). Let $B_e(x)$ be an open ball around $x \in \mathbb{R}^2$ with its radius $e > 0$. For any $\omega \in F$ and for $e > 0$, the tropical hyperplane defined by $\omega$ and $e$, denoted by $H_{\omega,e}$, is the set of points $f \in F$ such that
$$\{\omega(y) + f(y) = \omega(x^*) + f(x^*) : y \in (\mathbb{R}^2 - B_e(x^*))\} \geq 1,$$
where
$$x^* \in \arg\max_{x \in \mathbb{R}^2} (f(x) + \omega(x)).$$

We call $\omega \in F$ the normal vector of $H_{\omega,e}$.

Definition 6.7 (Tropical Distance to a Tropical Hyperplane in $F$). The tropical distance from a point $f \in F$ to a tropical hyperplane $H_{\omega,e}$ is defined as
$$d_{it}(f, H_{\omega,e}) := \min\{d_{it}(f, g) : g \in H_{\omega,e}\}.$$

Lemma 6.8.
$$d_{it}(f, H_{\omega,e}) = d_{it}(f + \omega, H_{\omega,0}).$$

Proof. $H_{\omega,e}$ is defined as:
$$H_{\omega,e} = \{g \in F : \{\omega(y) + g(y) = \omega(x^*) + g(x^*) : y \in (\mathbb{R}^2 - B_e(x^*))\} \geq 1\},$$
where
$$x^* \in \arg\max_{x \in \mathbb{R}^2} (\omega(x) + g(x)).$$

Thus
$$d_{it}(f, H_{\omega,e}) = \left\{d_{it}(f, g) : \{\omega(y) + g(y) = \omega(x^*) + g(x^*) : y \in (\mathbb{R}^2 - B_e(x^*))\} \geq 1\right\}.$$

Lemma 6.9. Let $x^* \in \arg\max_{x \in \mathbb{R}^2} f(x)$. Then $d_{it}(f, H_{\omega,0}) = \max\{f(x) : x \in \mathbb{R}^2 - B_e(x^*)\} - \max\{f(y) : y \in (\mathbb{R}^2 - B_e(x^*))\}$.

Proof. Let $\alpha = \max\{f(x) : x \in \mathbb{R}^2 - B_e(x^*)\}$ and $\beta = \max\{f(y) : y \in (\mathbb{R}^2 - B_e(x^*))\}$. Then, $g(x) = f(x) + \alpha I(y)$, where $\alpha = \max\{f(x) : x \in \mathbb{R}^2 - B_e(x^*)\}$ and $I(y)$ is the indicator function. Notice that $g(x)$ is in $H_{\omega,0}$.
Example 6.10 (Tropical Hyperplanes in $F$). We will use the same set up as in Example 6.1 and Figure 11. Let $e = 1$. Then by Lemma 6.9, we have

$$d_t(F_1, H_{0, e}) = \max \{ F_1(x) : x \in \mathbb{R} \} - \min \{ F_1(y) : y \in (\mathbb{R} - B_e(x^*)) \},$$

where $x^* = \arg \max_{x \in \mathbb{R}} (F_1(x))$. Let $F_1(y) = F_1(0) = 1$.

Similarly,

$$d_t(F_3, H_{0, e}) = \max \{ F_3(x) : x \in \mathbb{R} \} - \min \{ F_3(y) : y \in (\mathbb{R} - B_e(x^*)) \},$$

where $x^* = \arg \max_{x \in \mathbb{R}} (F_3(x))$.

Suppose we have $\omega = F_3$. Then, by Lemma 6.8, we have

$$d_t(F_1, H_{0, e}) = d_t(F_1 + \omega, H_{0, e}) = d_t(F_1, H_{0, e}) + d_t(\omega, H_{0, e}) = 0,$$

which is a Gaussian distribution function with $\mathbb{R}$. Suppose $\omega = 0$.

Example 6.12. We will use the same set up as in Example 6.1. Again, let $L$ be a set of univariate Gaussian distribution functions with $\mu$ and $\sigma$, and mixtures of these Gaussian distributions with real valued coefficients. Suppose $\omega = 0$ which is a Gaussian distribution function with $\mu = 0$ and $\sigma \rightarrow \infty$, $e = 1$, and $x = 0$. Then

$$S_{0, e}^0 := \{ \text{Mixtures of univariate Gaussian distributions with real coefficients whose argmax in } [-1, 1] \}.$$

Also note that a set of

$$\{ \text{Gaussian distribution functions with } \mu \in [-1, 1] \}$$

is in this open sector.

6.2. Tropical SVMs on Function Spaces

For a given $\epsilon > 0$, suppose $F \in F$ and $Y^1, Y^2 \in \{0, 1\}$ are random variables such that there exist $x^\ast \in F$ with

$$\left( \cup_{\omega \in \mathbb{R}} \max_{x \in \mathbb{R}} (x + \omega) + F_1(x^\ast) \right) \cap \left( \cup_{\omega \in \mathbb{R}} \max_{x \in \mathbb{R}} (x + \omega) + F_2(x^\ast) \right) = \emptyset. \quad (22)$$

Now we set up a tropical SVM over $F$, whose solution $\omega$ satisfies Equation 22. Let $D_F$ be the distribution on the joint random variable $(F, Y)$ for $F \in F$ and $Y \in \{0, 1\}$, and let $S_F$ be the sample $S_F := \{(F^1, Y^1), \ldots, (F^n, Y^n)\}$. For a given $\epsilon > 0$, we formulate an optimization problem for solving the normal vector $\omega \in F$ of an optimal tropical separating hyperplane $H_{0, \epsilon}$ for random variables $X \in F$ given $Y \in \{0, 1\}$: For some cost $C \subseteq \mathbb{R}$,

$$\min_{\omega \in \mathbb{R}} \max_{x \in \mathbb{R}} \left( \max_{\omega \in \mathbb{R}} (x + \omega) + F_1(x^\ast) \right)$$. \quad (23)

where $X^\ast \in \arg \max_{x \in \mathbb{R}} (\omega(x) + F(x))$ and $\omega \in \mathbb{R}$. In practice, we approximate each function $F_i$, for $i = 1, \ldots, n$, by its empirical function $\hat{F}_i$ by taking some $x \in \mathbb{R}^d$ to evaluate $F(x)$. In this paper we propose Algorithm 1 to heuristically conduct a tropical SVM using empirical functions $\hat{F}_i$, for $i = 1, \ldots, n$, using finite set of points $\{x^1, \ldots, x^k\} \subseteq \mathbb{R}^d$ with $y \notin B_i(x^j)$ for $j \neq i$ and for $i = 1, \ldots, k$. Let $\hat{F}_i = (F_i(x^1), \ldots, F_i(x^k))$, for $i = 1, \ldots, n$.

**Algorithm 1:** Heuristic tropical SVM over $F$

Input: A train set $S = \{(\hat{F}^1, Y^1), \ldots, (\hat{F}^n, Y^n)\}$

Output: Estimated normal vector $\hat{\omega}$ of a tropical hyperplane

Apply $\{(\hat{F}^1, Y^1), \ldots, (\hat{F}^n, Y^n)\}$ to a tropical SVM over $\mathbb{R}^d$.\textbackslash \mathbb{R}^1$

**return** The output from the tropical SVM over $\mathbb{R}^d$\textbackslash \mathbb{R}^1$

7. Discussion

We show the generalization error bounds for tropical SVMs over the tropical projection space $\mathbb{R}^d$ which is isometric to $\mathbb{R}^{d-1}$. These bounds still depend on the dimension $d$ and if we fix the sample size $n$ these bounds do not make sense and we cannot extend these bounds for tropical SVMs over a function space $F$ with tropical metric $d_t$. For future work it is interesting to obtain generalization error bounds for tropical SVMs over a function space $F$ with tropical metric $d_t$. In addition, computational experiments show that tropical SVMs over $\mathbb{R}^d$\textbackslash \mathbb{R}^1$ have much lower error rates than ones of $L_2$ norm SVMs over $\mathbb{R}^d$ when we fix the sample size and grow the dimension $d$. It seems these error rates are bounded by some constant. We are interested in tighter generalization error bounds for tropical SVMs over the tropical projection space $\mathbb{R}^d$\textbackslash \mathbb{R}^1$. The generalization error bound for tropical SVMs was derived without any assumptions on data distributions. Specifically, the bound was derived purely combinatorially only based on the shapes of the hyperplanes. Algorithmic details of tropical SVMs are not yet taken into account. Thus it is not surprising that the tropical SVMs under the max-plus algebra outperform in the case where a few axes are much more informative than the others. Thus the evaluation of the tropical SVM in computational experiments gave complementary information.
The max-plus algebra also leads to the anomalous $\log n$ scaling behaviors in the tropical distance $d_{\text{tr}}$ between random vectors of length $n$ and its scaling behavior is the direct consequence of the extreme value statistics. As it is also essential for the theoretical explanation of the robustness of tropical SVMs against the curse of dimensionality, extreme value statistics play the key role in the computation with the tropical metric.

For computational experiments, we used software developed by Tang et al. (2020). Tang et al. (2020) developed heuristic methods for hard margin and soft margin tropical SVMs over $\mathbb{R}^d/\mathbb{R}$ since it is infeasible to compute them in practice. We conjecture that computing optimal separating hyperplanes for tropical SVMs is NP-hard. However, we still do not know the exact computational time complexity in terms of $n$ and $d$.

References
Améndola, C., Klüppelberg, C., Lauritzen, S., Tran, N., 2020. Conditional independence in max-linear bayesian networks. Available at https://arxiv.org/abs/2002.09233.

Bartlett, P.L., Mendelson, S., 2003. Rademacher and gaussian complexities: Risk bounds and structural results. J. Mach. Learn. Res. 3, 463–482.

Boser, B.E., Guyon, I.M., Vapnik, V.N., 1992. A training algorithm for optimal margin classifiers, in: Proceedings of the Fifth Annual ACM Workshop on Computational Learning Theory, ACM Press. pp. 144–152.

Bradley, P.S., Mangasarian, O.L., 1998. Feature selection via concave minimization and support vector machines, in: Proceedings of the Fifteenth International Conference on Machine Learning, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA. p. 82–90.

Burges, J.C., 1998. A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery 2, 121–167.

Cohen, J.Y., Haesler, S., Vogl, L., Lowell, B.B., Uchida, N., 2012. Neurontype-specific signals for reward and punishment in the ventral tegmental area. Nature 482, 85–8.

Cortes, C., Vapnik, V., 1995. Support-vector networks, in: Machine Learning, pp. 273–297.

Dayan, P., Abbott, L.F., 2001. Theoretical Neuroscience. MIT Press.

Fan, J., Ren, Y., 2006. Statistical analysis of dna microarray data in infinite dimensional langevin dynamics, in: Larochelle, H., Ranzato, M., Jaderberg, W., Schmidhuber, J., Bottou, L. (Eds.), Advances in Neural Information Processing Systems, Curran Associates, Inc. URL: https://proceedings.neurips.cc/paper/2020/file/df1a336d7e8bcb8186de6680c43a9-Paper.pdf.

Gärnér, B., Jaggi, M., 2006. Tropical support vector machines. ACS Technical Report. No.: ACS-TR-362502-01.

Guerrmeur, Y., 2007. How sure are you that large margin implies low vc dimension. Journal of Machine Learning Research 8, 2551–2594.

Gumbel, E.J., 2004. Statistics of Extremes. Dover.

Hayashi, N., Watanabe, S., 2017. Upper bound of bayesian generalization error in non-negative matrix factorization. Neurocomputing 266, 21–28.

Heng-Chih Chiang, Moses, R.L., Potter, L.C., 2000. Model-classification of radar images. IEEE Transactions on Information Theory 46, 1842–1854. doi:10.1109/18.857795.

Ishikawa, T., Matsumoto, H., Miura, K., 2018. Discriminating dopamine from non-dopamine neurons based on the spontaneous electrical activities. Proc. of IEEE Brain Initiative Workshop on Advanced NeuroTechnologies.

Ishikawa, T., Matsumoto, H., Miura, K., 2019. Identification of midbrain dopamine neurons using features from spontaneous spike activity patterns. IEEE EMBC 2019, 2990–2993.

Jaggi, M., Katz, G., Wagner, U., 2008. New results in tropical discrete geometry.

Joswig, M., 2005. Tropical halfspaces. Combinatorial and computational geometry 52, 409–431.