Comparation Between Linear and Polynomial Kernel Function For Ovarium Cancer Classification

R S Samosir$^{1,2}$, F I. Gaol$^2$, B S Abbas$^2$, B S Sabarguna$^2$, and Y Heryadi$^2$

$^1$Institut Teknologi dan Bisnis Kalbis, Pulomas, Jakarta Timur, Indonesia
$^2$Bina Nusantara University, Kebon Jeruk, Jakarta Barat, Indonesia

ridha.samosir@kalbis.ac.id

Abstract. Ovarium cancer is a kind of cancer disease which often attacks a woman. Ovarium cancer consists of benign and malignant type. Based on the data set obtained, this research try to classify the data set with support vector machine algorithm. The principle of SVM is how to identify a hyperplanes with maximal margin. Hyperplanes is a border line between the data. Here, we will used two kernel function in SVM, linear and polynomial kernel function. Next We will compare both of them algorithm to reach the hyperplanes with maximal margin. The hyperplanes then used to predict the class of data testing given. Finally, comparison result will be obtained from confusion matrix and accuracy calculation for each kernel function while used to classify the ovarium cancer data set. Accuracy result with Linear kernel function is 68.98% for first schema and 66.67% for second schema. Then accuracy result with Polynomial Kernel Function is 83.79% for both of schema.

1. Introduction
Ovarium cancer is one kind of cancer that often attacks a woman. Ovarium cancer means the cancer cell in ovary growth fast. Ovary is a place where ovum and female hormones (estrogen and progesterone) is female reproductive glands in which the ova, or eggs are formed) is produced (estrogen and progesterone). Every woman have two ovaries, one on each side of the uterus. The ovaries each about the size of an almond. Ovarium cancer usually attacks middle or old woman. Generally. Cancer occur because there has been genetical change in human body that make normal cell growth into cancer cell. The things that increase risk of ovarium cancer are small number of offspring, a woman with breast cancer history or her family, elderly woman and woman who have had estrogen replace therapy. Most ovarium cancers are either ovarium epithelial cancers or malignant germ cell tumor. Ovarium cancer often goes undetected until it has spread. Some methods to diagnose ovarium cancer are CT Scan, Magnetic Resonance Imaging (MRI), Biopsi, and ultrasonography (USG). MRI more sensitive than CT Scan results. MRI results show that MRI able to detect location and size change of tumor. MRI results is read by radiologist through observation and analysis. Then that radiolog have a discussion with the doctor. Quality of MRI results determined how fast and accurate the diagnose. Fast refers to radiolog skills reads the MRI.

Classification is a supervised learning process. Supervised means there is a label or class for whole of data set. Classification is a technique for merge all similar data set into known class or group. Known class or label will be the supervision target for all of data set. Many algorithm for
classification, one of this is support vector machine (SVM). SVM has been widely applied to various real problems everyday such as medical, financial, other fields. First scientist for support vector machine is Vapnik in 1995. The principle of SVM is determine the hyperplane for whole data set. This hyperplanes claimed as classifier for each data set into class or label. The output given is the prediction class for the data testing given. The advantage of SVM is the capability of this algorithm to identify hyperplanes with maximal margin.

Mathematically, real problems in daily life grouped into linear and non linear problems. Although almost real problem refers to non linear problem include medical fields. Linear problem will be easily solved then non linear problem. Its because linear problem always consist homogen or structured data. Homogeneity facilitates the classification process. For non linear problems, SVM solve it with kernel function. Kernel function. Kernel functions aim to map x in input space into feature space F with higher dimension. In non linear problem, data can not separated linearly but can be separated on the feature space. Some of kernel function that usually used in SVM are linear kernel, polynomial kernel function, radial basic function, sigmoid function and multi layer perceptron. Technique to choose exactly kernel function is depend on the data set.

This research aims to compare between both of kernel function in SVM, linear kernel and polynomial kernel. Comparison between both of kernel function calculated with accuracy measurement between them when its used to classify the ovarium cancer data set. Classification aims to predict the class or label for data testing given. Class or label for data set defined are normal and cancer. Its mean classification with SVM in this research aims to predict the class or label for data testing given. Finally, we can calculate how many normal or cancer patient. Ovarium cancer data set for training used for this experiment is taken from Matlab Software. Size of data is 216 x 4000. All data set is classified into two class label. 216 Raws of data is representation of the number of patient. From 216 patient, there is 95 normal patient and 121 cancer patient. Column number is representation of ion intensity level in ovaries cell. Than for data testing, this experiment is taken from data training. There is 151 raws of data testing.

2. Literature Review

In this section will explain support vector machine, feature extraction, principle component analysis, polynomial kernel function, linear kernel function and confusion matrix.

2.1. Support Vector Machine

Support vector machine or SVD is has widely applied for any real problems in our daily life such as medival, financial and etc. The principle of SVM is to determine hyperplanes for all data set. The hyperplanes is called as classifier because hyperplanes aims to separate all data set into their class or label. To separated the data, we can use linear and non linear technique. Kernel is a function that map all feature in data set from the first dimension into higher dimension feature. Kernel function approach is one uniq methods that make SVM is different with others classification algorithm. Non linear kernel function have many approach such as polynomial, radial basis function, quadratic or sigmoid and multi layer perceptron. Technique to choose kernel function will be affect to classifier or hyperplanes.

The output from classification process with SVM is class or laber for new data testing. The advantage of SVM algorithm is capability of SVM to determine hyperplaness with maximal marginal. Basically, figure 2 shows classification concept with SVM and show how the hyperplanes separate data. Figure 1 and figure 2 show example between linear and non linear data separated with hyperplaness.
2.2. Feature Extraction

Feature extraction is a process to extract the feature from image that will be analysis for the next step. Feature represents the characteristic of image. For example is flower image. From that image then we can identify all the feature such as pethal, sephal and others. Two type of feature are natural or artificial. Natural feature embedded with the image such as brightness, edge texture and other. But artificial feature shows some operation technique to the image. Characteristic from feature extraction then will be used as parameter or input to see the different among the object before used in next steps. Feature extraction will always be needed even while we have high dimension data set. Generally, extraction feature try to extract texture, colour, contour, size and others.

Principal component analysis or PCA is one of feature extraction methods. PCA will reduces some independent features. PCA usually used for data compression. PCA was invented by Karl Pearson. The principle of PCA is do transformation the size of data. PCA choose the non redundant and irrelevant data, its mean that type of data above have important rules as original data variety. Other words, PCA is a dimension reduction tool through feature reduction but not throw the important
information from the data. PCA select a subset of feature from large data set, based on which feature have a correlation with principal component. Principal component is new feature selected.

2.3. Linear Kernel
Hyperplanes determination for the linear data step problems will be easy to reach. Hyperplanes will be asy to reach. Its because, basically SVM have capability work for linear separated data. According to figure 2 then linear function will be easy to separated all linear data. Linear data means its has high homogeneity. Other words say data set with lower dimension. Here is the mathematics formul for linear kernel function:

\[ K(x,y) = x \cdot y \]  

(1)

2.4. Polynomial Kernel
Actually, Polynomial kernel is one of kernel transformation on SVM for non linear problem. Based on explanation above, non linear kernel function will be fit for heterogen data. Heterogen data means data with non linear distribution. Non linear data distribution will be difficult if solved by linear kernel. Because we can not determined the hyperplanes for non linear data easily. Kernel is a function that map data from lower dimension to higher dimension. Other words, its map X in input space to feature space. This Kernel approach is the difference SVM with other classification algorithm. Usually, other classification algorithm try to reduce the dimension of data or make it simply. Reduce dimension aims to make computation process more simply.

2.5. Confussion Matrix
Measure the performance of classification algorithm show how better that algorithm can classify the data testing given. Confussion matrix can represents information as the classification result. Result from confussion matrix consist of 4 variable, there are true positive (TP), true negative (TN), false positive (FP) and false negative (FN). True positive identify how many positive data that result correct identified by the system. True negative identify how many negative data that result correct identified by system. False positive means number of data positive that wrong result identified by system. And false negative means number of negatif data that represents wrong result identified by system. Based on that 4 value so we can identify number of accuration, precision and recall. Here is the mathematic formula to looking for accuration number according to four value in confussion matrix.

3. Research Methodology
In this section will explain about research object, research method used and research step.

3.1. Research Methods
To obtain data, research team load from Matlab Software. Matlab provide some data for experiment. Before do the classification, research team do some pre processing to clear data from empty record, inmatch data format and other. In addition, pre process with feature extraction is done to reduce computation time while learning process. For classification, team will use SVM model and compare between linear kernel and polynomial kernel in SVM model. Finally, we will see percentage accuracy each kernel function. Accuration obtained from calculation of confussion matrix.

3.2. Research Object
According to the previous explanation, the object of this research is 216 x 4000 ovarium cancer data set. From 216 rows, data training indicate 95 normal and 121 cancer. 4000 show the column number of data. Number of column represent of ion intensity level. Ion intensity as an indicator whether a person is attacked by ovarium cancer or not.
3.3. Research Step
Some step has done for this research is started with identify the data set and classification algorithm will used. Next stages is design the program architecture. Program architecture will help us when develop the system. Program architecture show all methods and object used to develop the system. Then we develop the system. System is developed using Matlab Software Versi R2014a. In the sytem we declare both of kernel function used, because objective of this research is comparation between linear kernel and polynomial kernel. Last stages is identify the percentage accuration each function. Percentage accuration is calculated using confusion matrix.

4. Result and Discussion
Here will be showed all experiment result has done using software Matlab: dataset plotting result, Feature Extraction result, classification with polynomial kernel function result and classification with linear kernel function result.

4.1. Plotting Data Set
Figure 3 show how data training plotting. According to explanantion above show that ovarium cancer data set with 216x4000 size is plotted in input space. From that image above, the red one is cancer distibuted data while the green one show normal distributed data. If we look at that image, data set is distributed in non linear form because it will be difficult to determine the hyperplanes. But for this experiment research, we try to solve that with linear and non linear kernel. Then we can compare both of them from their each accuration number.

4.2. Result Of Feature Extraction
From the size of data set show that there is 4000 column of data. Number of column is representation of ion intensity level. For reduce some complexcity number of computation process, we need a feature to extrac the fit feature according to data. In classification algorithm, we call it Feature extraction. In this experiment, we propose to use PCA (Principal Component Analysis) for feature extraction methods. PCA choosing because PCA fit to two dimension plotting, so from many feature have according to data then we choose two best feature. Best feature means feature which have large participation in data. With PCA, we can choose two maximal feature.
4.3. Classification Result With Linear Kernel Function
After feature extraction is done then do classification process. Figure 4 show classification result for data testing given using linear kernel function.

![Figure 4](image)

**Figure 4.** Classification Result With Linear Kernel Function.
This figure show distribution of data after classification process using linear kernel function in SVM approach.

From that picture above show that the hyperplanes separate the the data set to be normal and cancer. The purple and blue points shows classification result. Purple points represents cancer classified and blue points represents normal classified. We can see that some green points occur in purple side. Green points is data training for normal class. So do the other hands, its show that there is some red points in the blue side. Its mean linear kernel function can not classify all of data set fit to each class.

4.4. Classification Result With Polynomial Kernel Function
Figure 5 show classification result for data testing given using linear kernel function. According to the explanation above that non linear kernel function give better classification result while the data distribution is not linear or not homogen. Non linear kernel function like polynomial can determine the hyperplanes in high dimension or feature space. We can see from picture above that polynomial as classifier give better result to classify all of data testing. Although some points is not classied well fit to each class, but number of point is few than the first methods. All cancer points is classified fit to each class. Then we can see there is more than one hyperplaness showed, because polynomial work in feature space not in input space.

![Figure 5](image)

**Figure 5.** Classification Result With Polynomial Kernel Function. This figure show distribution of data after
classification process using polynomial kernel function in SVM approach.

4.5. Accuration

To identify comparison between linear and polynomial, we use accuration number. Accuration number is obtained from some experiment with data training and data testing. Before we identify the accuration, we try some schema for data training and data testing. Table 1 show some schema for data training and data testing.

| No | Kernel Function | Data Training | Data Testing |
|----|-----------------|---------------|--------------|
| 1  | Linear Kernel Function | xdata=obs(1:200,30:70) | data_testing=obs(20:200,30:70) |
|    |                  | xdata=obs(1:end,30:70) | data_testing=obs(20:200,30:70) |
| 2  | Polynomial Kernel Function | xdata=obs(1:200,30:70) | data_testing=obs(20:200,30:70) |
|    |                  | xdata=obs(1:end,30:70) | data_testing=obs(20:200,30:70) |

For each kernel function used, we do experiment with two schema of data training and one schema of data testing. Two schema of data training is raw 1 to 160 and all of raw data used. Then for data testing we take from data training, raw 50 to raw 200. All of data training and data testing schema we execute with kernel function used, linear and polynomial. Table 2 show accuration result include confusion matrix each kernel function used. Based on that picture, we can see that polynomial give higher accuration than linear kernel. From the first picture, we can see that data set has non linear distribution. So thats why polynomial higher in accuration. If we look further, there is little difference in polynomial kernel function according two schema of data set. Second schema give higher accuration than the first one. Second schema involve all raw of data set. Just the same with linear kernel, first schema give high accuration number.

| Kernel Function | Confusion Matrix | Accuration |
|-----------------|------------------|------------|
| Linear Kernel Function | 85 15 17 64 78 13 24 66 | 68,98% 66,67 % |
| Polynomial Kernel Function | 102 0 102 0 | 83,79% |
|                      | 0 79 0 79 |            |
5. Conclusion

Based on all experiment has done, we take some resume here:

1. Polynomial kernel function gives higher accuracy than linear kernel. Its because from data set, we can see there is non linear data set distribution.
2. Classification result show better while classifier involve all data set. Because quantity of data set give affect to learning process.
3. Both of kernel used show that number of feature for training data affect to accuracy. If we involve too few or too many feature then the accuracy will be lower. But if we involve fix feature then accuracy number will be increase.
4. While we choose data testing, its better to use outside of data training. Its because we need new data to test our classifier. If we used data testing same with data training, it will give high accuracy because all data has been learned

We hope this experiment can use for other research, then here some recommendation according to this research:

1. Selection of kernel function affect to classification result. Its because each kernel function determine the fit hyperplanes according to data set.
2. Data training schema affect to accuracy. In SVM, more data involved then better in accuracy. In SVM number of data set helps the learning process.
3. Number of feature to train can be considered to increase accuracy result.

Last but not the least, we can consider to use cross validation to choose most appropriate kernel function.

References

[1] https://www.mayoclinic.org/diseases-conditions/ovarium-cancer, access at January 2017
[2] Usman K and Kaship R 2017 Brain Tumor Classification From Multi Modality MRI Using Wavelets and Machine Learning Pattern Analytics Application. 20 3 871
[3] Samosir R S 2017 Filtering and Wavelet Transform Algorithm For Old Document Image Restoration Computer, Mathematics and Engineering Application. 8 3 177-181
[4] Megha P A and Ram Mohana Reddy G 2013 An intelligent Content-Based Image Retrieval system For Clinical Decision Support in Brain Tumor Diagnosis International Journal Multimed Information Retrieval. 2 3 175-188
[5] Taranjit K, Barjinder S S and Savita G 2018 A Novel Feature Selection Method For Brain Tumor MR Image Classification Based on The Fisher Criterion and Parameter-Free Bat Optimization Neural Computing & Application. 29 8 193-206
[6] Chaddad A 2015 Automated Feature Extraction in Brain Tumor by Magnetic Resonance Image Using Gaussian Mixture Models J. Biomed Imaging.
[7] Zongmei T, Nilanjan D et al 2017 Morphological Segmenting and Neighborhood Pixel-Based Locality Preserving Projection on Brain fMRI Dataset for Semantic Feature Extraction: an effective computing study Neural Comput & Application.
[8] Christianini N and Shawe-Taylor J 2000 An Introduction to Support Vector Machine and Other Kernel Based Learning Methods Linear Kernel (Cambridge United Kingdom : Cambridge Univeristy Press)
[9] Johan A K S, Tony V G et al 2000 Least Square Support Vector Machine (London : World Scientific Publishing)
[10] Christoper D M, Prabhakar R and Hinrich S 2009 Introduction to Information Retrieval (England: online Edition Cambridge University Press)
[11] Jingxiang Z and Shitong W 2016 A Fast Leave-One-Out Cross Validation for SVM Like Family Neural Computing and Applications. 27 6 1717-30