Fixing an error in Caponnetto and de Vito (2007)

Dougal J. Sutherland
Gatsby unit, University College London
dougal@gmail.com
February 2017

Abstract
The seminal paper of Caponnetto and de Vito (2007) provides minimax-optimal rates for kernel ridge regression in a very general setting. Its proof, however, contains an error in its bound on the effective dimensionality. In this note, we explain the mistake, provide a correct bound, and show that the main theorem remains true.

The mistake lies in Proposition 3’s bound on the effective dimensionality $\mathcal{N}(\lambda)$, particularly its dependence on the parameters of the family of distributions $b$ and $\beta$. We discuss the mistake and provide a correct bound in Section 1. Its dependence on the regularization parameter $\lambda$, however, was correct, so the proof of Theorem 1 carries through with the exact same strategy. The proof was written in such a way, though, that it is not immediately obvious that it still holds for the corrected bound; we thus provide a more detailed explication of the proof, showing it is still valid.

This note will make little sense without a copy of the original paper at hand. Numbered theorems and equation references always refer to those of Caponnetto and de Vito (2007); equations in this document are labeled alphabetically.

A trivial correction First, we note a tiny mistake: Theorem 4 needs $C_\eta = 96 \log^2 \frac{\eta}{n}$, rather than $32 \log^2 \frac{\eta}{n}$, because the last line of its proof dropped the constant 3 in front of $S_1(\lambda, \mathbf{z})$ and $S_2(\lambda, \mathbf{z})$ in (36).

1 Bound on the effective dimensionality
Part of Proposition 3 is the claim that for $p \in \mathcal{P}(b, c)$, with $c \in [1, 2]$ and $b \in (1, \infty),$

$$\mathcal{N}(\lambda) \leq \frac{\beta b}{b-1} \lambda^{-\frac{\beta}{b}}. \quad (a)$$

The argument starts like this:

$$\mathcal{N}(\lambda) = \text{Tr} \left[ (T + \lambda I)^{-1} T \right]$$
$$= \sum_{n=1}^{\infty} \frac{t_n}{t_n + \lambda}$$
$$\leq \sum_{n=1}^{\infty} \frac{\beta/n^b}{\beta/n^b + \lambda} = \sum_{n=1}^{\infty} \frac{\beta}{\beta + \lambda n^b} \quad (b)$$
$$\leq \int_{0}^{\infty} \frac{\beta}{\beta + \lambda x^b} \, dx$$
$$= \lambda^{-\frac{1}{b}} \int_{0}^{\infty} \frac{\beta}{\beta + \tau^b} \, d\tau, \quad (c)$$
with (b) following from Definition 1 (iii), the next line’s upper bound by an integral true since \(x \mapsto \frac{\beta}{\beta + \lambda x^\tau}\) is decreasing, and then doing a change of variables to \(\tau^b = \lambda x^b\).

But then the paper claims without further justification that

\[
\int_0^\infty \frac{1}{\beta + \tau^b} d\tau \leq \frac{b}{b - 1}.
\]  

(d)

In fact, (d) is incorrect: as \(\beta \to 0\), the integral approaches the divergent integral \(\int_0^\infty \tau^{-b} d\tau\), but \(\frac{b}{b - 1}\) clearly does not depend on \(\beta\). We can instead compute the true value of the integral:

\[
\int_0^\infty \frac{1}{\beta + \tau^b} d\tau = \int_0^\infty \frac{1/\beta}{1 + \left(\tau^{\beta^{-1}}\right)^b} d\tau \\
= \beta^{-1} \int_0^\infty \frac{1}{1 + u^\beta} \beta^{1/\beta} du \\
= \beta^{1+\beta} \frac{\pi/b}{\sin(\pi/b)}.
\]  

(e)

Using (e) in (c), we get a correct lower bound:

\[
\mathcal{N}(\lambda) \leq \beta^{1+\beta} \frac{\pi/b}{\sin(\pi/b)} \lambda^{-1/\beta}.
\]  

(f)

(f) has the same dependence on \(\lambda\) as (a), but the dependence on \(\beta\) and \(b\) differs.

To demonstrate, we now plot the sum (b) (green, middle), the correct upper bound (f) (blue, top), and the purported upper bound (a) obtained via (d) (orange, bottom) for \(\beta = 0.1, \lambda = 10^{-3}\).

2 Consequences in Theorem 1

Though it isn’t obvious at first, the proof of Theorem 1 depends on the \(\mathcal{N}(\lambda)\) bound only in its rate on \(\lambda\), which was indeed correct: thus the proof of Theorem 1 remains valid. We now restate the relevant parts of the proof of Theorem 1 in a way that makes this lack of dependence more explicit.
Theorem 4 gives us that for any any $\eta \in (0, 1)$, with probability greater than $1 - \eta$ we have
\[
\mathcal{E}[f^\lambda] - \mathcal{E}[f_{\mathcal{H}}] \leq C_\eta \left( A(\lambda) + \frac{\kappa^2}{\ell^2 \lambda^3} B(\lambda) + \frac{\kappa}{\ell \lambda} A(\lambda) + \frac{\kappa M^2}{\ell^2 \lambda} + \frac{\Sigma^2}{\ell} N(\lambda) \right),
\]
provided that
\[
\ell \geq \frac{2C_\eta \kappa}{\lambda} N(\lambda) \quad \text{and} \quad \lambda \leq \|T\|_{\mathcal{L}(\mathcal{H})}.
\]

Define $Q$ as
\[
Q = \begin{cases} 
\beta \frac{\pi b}{\sin(\pi b)} & b < \infty, \\
\beta & b = \infty,
\end{cases}
\]
so that, from Proposition 3 and (f),
\[
A(\lambda) \leq \lambda^c \|T^{\frac{1-\epsilon}{2}} f_{\mathcal{H}}\|^2_H, \quad B(\lambda) \leq \lambda^{c-1} \|T^{\frac{1-\epsilon}{2}} f_{\mathcal{H}}\|^2_H, \quad N(\lambda) \leq Q \lambda^{-\frac{1}{2}}.
\]

Plugging in these rates and $\|T^{\frac{1-\epsilon}{2}} f_{\mathcal{H}}\|^2_H \leq R$ from Definition 1 (ii), we have that
\[
\mathcal{E}[f^\lambda] - \mathcal{E}[f_{\mathcal{H}}] \leq C_\eta \left( R\lambda^c + \kappa R \lambda^{c-2} + \frac{\kappa R}{\ell} \lambda^{-1} + \kappa M^2 \frac{\lambda^{-1}}{\ell} + \Sigma^2 Q \lambda^{-\frac{1}{2}} \right) \leq C_\eta \lambda^{-\frac{1}{2}}
\]
provided that
\[
\ell \geq 2C_\eta Q \lambda^{-\frac{1}{2} - \frac{b+1}{b}}.
\]

Note that for $b = \infty$, nothing has changed from the paper. Thus the proofs (which were not written explicitly in the paper) remain the same. We thus assume $b < \infty$.

2.1 $c > 1$

When $c > 1$, let
\[
\ell \eta \geq (2C_\eta Q)^{\frac{b+1}{b}}
\]
so that for any $\ell \geq \ell \eta$ we have
\[
\frac{\ell}{\frac{b}{b+1}} \geq 2C_\eta Q.
\]

Define $\lambda_\ell = \ell^{-\frac{b}{b+1}}$. Then
\[
2C_\eta Q \lambda_\ell^{b+1} = 2C_\eta Q \left( \ell^{-\frac{b}{b+1}} \right)^{-\frac{b+1}{b}} = 2C_\eta Q \ell \frac{b+1}{b+1} = 2C_\eta Q \ell \frac{b+1}{b+1} \frac{b(c-1)}{b+1} \frac{b(c-1)}{b+1} \leq \ell \frac{b+1}{b+1} \frac{b(c-1)}{b+1} \frac{b(c-1)}{b+1} = \ell,
\]
so that (h) holds for $\lambda = \lambda_\ell$, and therefore (g) holds with probability at least $1 - \eta$ as long as $\lambda_\ell \leq \|T\|_{\mathcal{L}(\mathcal{H})}$. By Definition 1 (iii), the latter is at least $\lambda$; thus this condition is met as long as
\[
\lambda_\ell = \ell^{-\frac{b}{b+1}} \leq \alpha \quad \text{i.e.} \quad \ell \geq \alpha^{-\frac{b+1}{b}}.
\]
We thus don’t have to worry about it in the asymptotics. Plugging $\lambda_\ell$ into (g), we get
\[
\mathcal{E}[f^\lambda] - \mathcal{E}[f_{\mathcal{H}}] \leq C_\eta \left( R\ell^{-\frac{b}{b+1}} + \kappa R \ell^{-\frac{b(c-1)}{b+1}} + \kappa R \ell^{-\frac{b(c-1)}{b+1}} + \kappa M^2 \ell^{-\frac{b(c-1)}{b+1}} + \Sigma^2 Q \ell^{-\frac{b}{b+1}} \right)
\]

\[
= C_\eta \left( R\ell^{-\frac{b}{b+1}} + \kappa R \ell^{-\frac{b(c-1)}{b+1}} + \kappa R \ell^{-\frac{b(c-1)}{b+1}} + \kappa M^2 \ell^{-\frac{b(c-1)}{b+1}} + \Sigma^2 Q \ell^{-\frac{b}{b+1}} \right).
\]

3
Note that
\[3bc - 2b + 2 = bc + 2b(c - 1) + 2 > bc\]
\[2bc - b + 1 = bc + b(c - 1) + 1 > bc\]
\[2bc - b + 2 = bc + b(c - 1) + 2 > bc.\]
Thus for large \(\ell\) the \(\ell^{-\frac{\log \ell}{\ell}}\) terms dominate, and so we have that
\[\mathcal{E}[f_{\mathbf{z}}] - \mathcal{E}[f_{\mathcal{H}}] \leq 2C_\eta D\ell^{-\frac{\log \ell}{\ell}}\]
\[\forall \ell \geq \ell_\eta,\]
where \(D\) is some complex function of \(R, \kappa, M, \Sigma, \beta, b,\) and \(c\). Letting \(\tau = 2C_\eta D = 192D\log^2 \frac{6}{\eta}\) and solving for \(\eta\), we get \(\eta_r = 6e^{-\sqrt{\frac{\eta}{\tau}}}\). Thus
\[\Pr_{z \sim \rho^\ell} \left[ \mathcal{E}[f_{\mathbf{z}}] - \mathcal{E}[f_{\mathcal{H}}] > \tau \ell^{-\frac{\log \ell}{\ell}} \right] \leq \eta_r \quad \forall \ell \leq \ell_\eta,\]
and so
\[\limsup_{\ell \to \infty} \sup_{\rho \in \mathcal{P}(b,c,\rho^\ell)} \Pr_{z \sim \rho^\ell} \left[ \mathcal{E}[f_{\mathbf{z}}] - \mathcal{E}[f_{\mathcal{H}}] > \tau \ell^{-\frac{\log \ell}{\ell}} \right] \leq \eta_r,\]
and since \(\lim_{\tau \to \infty} \eta_r = \lim_{\tau \to \infty} 6e^{-\sqrt{\frac{\eta}{\tau}}} = 0\), we have as desired that
\[\lim_{\tau \to \infty} \limsup_{\ell \to \infty} \sup_{\rho \in \mathcal{P}(b,c,\rho^\ell)} \Pr_{z \sim \rho^\ell} \left[ \mathcal{E}[f_{\mathbf{z}}] - \mathcal{E}[f_{\mathcal{H}}] > \tau \ell^{-\frac{\log \ell}{\ell}} \right] = 0.\]

2.1.1 \(c = 1\)

Here we define \(\lambda_\ell = \left(\frac{\log \ell}{\ell}\right)^{\frac{1}{\ell^2}}\), so that the \(\ell\) requirement of (h) is
\[\ell \geq 2C_\eta \kappa Q \frac{\ell}{\log \ell},\]
that is,
\[\ell \geq \exp(2C_\eta \kappa Q),\]
so choosing \(\ell_\eta = \exp(2C_\eta \kappa Q)\) suffices.

As in the \(c > 1\) case, plugging \(\lambda_\ell\) into (g) obtains that as long as \(\ell \geq \ell_\eta\) (and \(\lambda_\ell \leq \alpha\)),
\[\mathcal{E}[f_{\mathbf{z}}] - \mathcal{E}[f_{\mathcal{H}}] \leq C_\eta \left( R\lambda_\ell + \kappa^2 R\frac{\lambda_\ell^{-1}}{\ell^2} + \kappa R\frac{1}{\ell^2} + \kappa M^2 \frac{\lambda_\ell^{-1}}{\ell^2} + \Sigma^2 Q\frac{\lambda_\ell^{-\frac{1}{2}}}{\ell} \right)\]
\[= C_\eta \left( R \left(\frac{\log \ell}{\ell}\right)^{\frac{1}{\ell^2}} + \kappa^2 R(\log \ell)^{\frac{1}{\ell^2}} \ell^{\frac{1}{\ell^2}} \ell^{-\frac{1}{\ell^2}} + \kappa R \ell^{-1} + \kappa M^2 (\log \ell)^{-\frac{1}{\ell^2}} \ell^{\frac{1}{\ell^2}} \ell^{-2} + \Sigma^2 Q(\log \ell)^{-\frac{1}{\ell^2}} \ell^{\frac{1}{\ell^2}} \ell^{-1}\right)\]
\[= C_\eta \left( R \left(\frac{\log \ell}{\ell}\right)^{\frac{1}{\ell^2}} + \kappa^2 R(\log \ell)^{\frac{1}{\ell^2}} \ell^{-\frac{1}{\ell^2}} + \kappa R \ell^{-1} + \kappa M^2 (\log \ell)^{-\frac{1}{\ell^2}} \ell^{-\frac{1}{\ell^2}} + \Sigma^2 Q(\log \ell)^{-\frac{1}{\ell^2}} \ell^{-\frac{1}{\ell^2}} \right).\]
The \(\ell^{-\frac{1}{\ell^2}}\) terms dominate, and so we can find a value \(D'\) such that
\[\mathcal{E}[f_{\mathbf{z}}] - \mathcal{E}[f_{\mathcal{H}}] \leq 2C_\eta D'\ell^{-\frac{1}{\ell^2}} \quad \forall \ell \geq \ell_\eta,\]
and the result follows by the same reasoning.

References
Caponnetto, A. and E. de Vito (2007). “Optimal Rates for the Regularized Least-Squares Algorithm”. In: Foundations of Computational Mathematics 7.3, pp. 331–368. doi: 10.1007/s10208-006-0196-8.