Stability of Tikhonov Regularization

Lorenzo Rosasco, Tomaso Poggio

9.520 Class 7

2010

About this class

Goal To show that Tikhonov regularization in RKHS satisfies a strong notion of stability, namely β-stability, so that we can derive generalization bounds using the results in the last class.

Plan

- Review of Generalization Bounds via Stability
- Stability of Tikhonov Regularization Algorithms

Learning algorithm and Generalization Error

A **learning algorithm** $\mathcal{A}$ is a map

$$S \mapsto f_S^\lambda$$

where $S = (x_1, y_1) \ldots . (x_n, y_n)$.

A **generalization bound** is a (probabilistic) bound on the defect (generalization error)

$$D[f_S^\lambda] = I[f_S^\lambda] - I_S[f_S^\lambda]$$

Uniform Stability

Let $S = \{z_1, ..., z_n\}$; $S^{i,z} = \{z_1, ..., z_{i-1}, z, z_{i+1}, ..., z_n\}$
An algorithm $\mathcal{A}$ is β-stable if

$$\forall (S, z) \in \mathcal{Z}^{n+1},\ \forall i,\ \sup_{z' \in \mathcal{Z}} |V(f_S^\lambda, z') - V(f_{S^{i,z}}^\lambda, z')| \le \beta.$$

Generalization Bounds Via Uniform Stability

From the last class we have that,

- If $\beta = \frac{k}{n}$ for some k,
- the loss is bounded by M,

then:

$$P\left(|I[f_S^\lambda] - I_S[f_S^\lambda]| \geq \frac{k}{n} + \epsilon\right) \leq 2\exp\left(-\frac{n\epsilon^2}{2(k+M)^2}\right).$$

Equivalently, with probability $1-\delta$,

$$I[f_S^\lambda] \leq I_S[f_S^\lambda] + \frac{k}{n} + (2k+M)\sqrt{\frac{2\ln(2/\delta)}{n}}.$$

β-stability of Tikhonov regularization

Today we prove that Tikhonov regularization

$$f_S^\lambda = \arg\min_{f \in \mathcal{H}} \{ \frac{1}{n} \sum_{i=1}^{n} V(f(x_i), y_i) + \lambda \|f\|_K^2 \}$$

satisfies

$$\forall (S, z) \in Z^{n+1},\ \forall i,\ \sup_{z' \in Z} |V(f_S^\lambda, z') - V(f_{S^{i,z}}^\lambda, z')| \leq \beta.$$

Preliminaries I

We assume the loss to be Lipschitz

$$|V(f_1(x), y') - V(f_2(x), y')| \leq L\|f_1 - f_2\|_\infty = L \sup_{x \in X} |f_1(x) - f_2(x)|$$

- The hinge loss and the ϵ-insensitive loss are both L-Lipschitz with $L = 1$ (exercise!).
- The square loss function is L Lipschitz if we can bound the values of y and $f(x)$.
- The $0 - 1$ loss function is not L-Lipschitz (why?)

Preliminaries II

If $f \in \mathcal{H}$ is in a RKHS with

$$\sup_{x \in X} K(x, x) \leq \kappa^2 < \infty$$

then

$$\|f\|_\infty \leq \kappa ||f||_K.$$

In particular this implies

$$\|f - f'\|_\infty \leq \kappa \|f - f'\|_K.$$

for any $f, f' \in \mathcal{H}$.

Preliminaries III

A key lemma

We will prove the following lemma about **Tikhonov regularization**:

$$||f_S^\lambda - f_{S^{i,z}}^\lambda||_K^2 \leq \frac{L\|f_S^\lambda - f_{S^{i,z}}^\lambda\|_\infty}{\lambda n}$$

This results is not straightforward and will be the most difficult part of the proof.

Proving Stability

1. assumption: $|V(f_1(x), y') - V(f_2(x), y')| \leq L\|f_1 - f_2\|_\infty$
2. property of RKHS: $\|f - f'\|_\infty \leq \kappa\|f - f'\|_K$, for any $f, f' \in \mathcal{H}$.
3. lemma: $||f_S^\lambda - f_{S^{i,z}}^\lambda||_K^2 \leq \frac{L\|f_S^\lambda - f_{S^{i,z}}^\lambda\|_\infty}{\lambda n}$

putting all together:

$$\begin{aligned}
|V(f_S^\lambda, z) - V(f_{S^{z,i}}^\lambda, z)| &\leq L\|f_S^\lambda - f_{S^{z,i}}^\lambda\|_\infty \\
&\leq L\kappa||f_S^\lambda - f_{S^{z,i}}^\lambda||_K \\
&\leq \frac{L^2\kappa^2}{\lambda n} \\
&=: \beta
\end{aligned}$$

Proving the Lemma

We now prove

$$||f_S^\lambda - f_{S^{i,z}}^\lambda||_K^2 \leq \frac{L\|f_S^\lambda - f_{S^{i,z}}^\lambda\|_\infty}{\lambda n}$$

Note that it holds only when we consider the minimizers of Tikhonov regularization.

We need again some preliminary facts and definitions...

Prelimaries: Derivative of a Functional

Let $F : \mathcal{H} \to \mathbb{R}$, f is differentiable at f_0 if

$$\lim_{t \to 0} \frac{F(f_0 + th) - F(f_0)}{t} = \langle \nabla F(f_0), h \rangle, \quad \forall h \in \mathcal{H}$$

and $\nabla F(f_0)$ is the called derivative.

Example: $F(f) = \|f\|^2 = \langle f, f \rangle$

$$\frac{\langle f_0 + th, f_0 + th \rangle - \langle f_0, f_0 \rangle}{t} = \frac{2t \langle f_0, h \rangle - t^2 \langle h, h \rangle}{t}$$

and taking $t \to 0$

$$\nabla F(f_0) = 2f_0.$$

Prelimaries: Bregman Divergence

Let $F : \mathcal{H} \to \mathbb{R}$ be a convex and differentiable function.

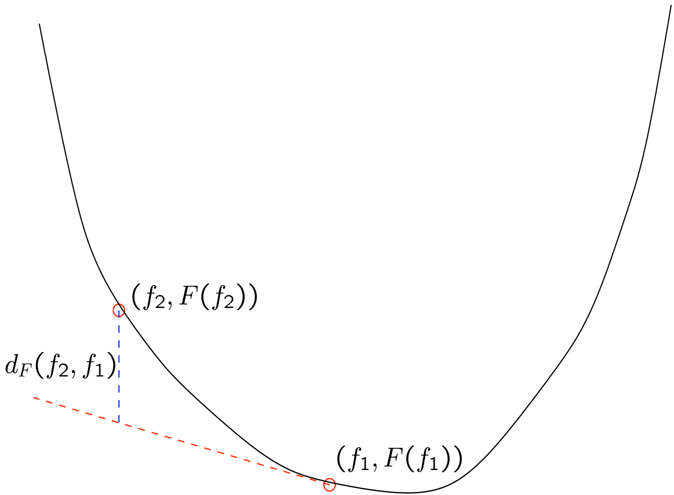
The Bregman divergence

$$d_F(f_2, f_1) = F(f_2) - F(f_1) - \langle f_2 - f_1, \nabla F(f_1) \rangle.$$

It can be seen as the error we make when we know $F(f_1)$ for some f_1 and "guess" $F(f_2)$ by considering a linear approximation to F at f_1:

$$F(f_2) = F(f_1) + \langle f_2 - f_1, \nabla F(f_1) \rangle.$$

Divergences Illustrated



Properties of Bregman Divergence

$$d_F(f_2, f_1) = F(f_2) - F(f_1) - \langle f_2 - f_1, \nabla F(f_1) \rangle.$$

We will need the following key facts about divergences:

- $d_F(f_2, f_1) \geq 0$
- If f_1 minimizes F, then the gradient is zero, and $d_F(f_2, f_1) = F(f_2) - F(f_1)$.
- If $F = A + B$, where A and B are also convex and differentiable, then $d_F(f_2, f_1) = d_A(f_2, f_1) + d_B(f_2, f_1)$ (derivative is additive).

The Tikhonov Functionals

We use the following short notation:

$$T_S(f) = \frac{1}{n}\sum_{i=1}^{n} V(f(x_i), y_i) + \lambda ||f||_K^2,$$

$$I_S(f) = \frac{1}{n}\sum_{i=1}^{n} V(f(x_i), y_i)$$

$$N(f) = ||f||_K^2.$$

Hence, $T_S(f) = I_S(f) + \lambda N(f)$. If the loss function is convex (in the first variable), then all three functionals are convex.

Proving the Lemma

We want to prove that

$$||f^{\lambda}_{S^{i,z}} - f^{\lambda}_{S}||^2_K \leq \frac{2L\|f^{\lambda}_{S} - f^{\lambda}_{S^{i,z}}\|_{\infty}}{\lambda n}$$

The proof consists of two steps:
Step 1: prove that

$$2||f^{\lambda}_{S^{i,z}} - f^{\lambda}_{S}||^2_K = d_N(f^{\lambda}_{S^{i,z}}, f^{\lambda}_{S}) + d_N(f^{\lambda}_{S}, f^{\lambda}_{S^{i,z}})$$

Step 2: prove that

$$d_N(f^{\lambda}_{S^{i,z}}, f^{\lambda}_{S}) + d_N(f^{\lambda}_{S}, f^{\lambda}_{S^{i,z}}) \leq \frac{2L\|f^{\lambda}_{S} - f^{\lambda}_{S^{i,z}}\|_{\infty}}{\lambda n}$$

Step 1

$$2||f^{\lambda}_{S^{i,z}} - f^{\lambda}_{S}||^2_K = d_N(f^{\lambda}_{S^{i,z}}, f^{\lambda}_{S}) + d_N(f^{\lambda}_{S}, f^{\lambda}_{S^{i,z}})$$

Recalling that $\nabla N(f) = 2f$, we have

$$\begin{aligned} d_N(f^{\lambda}_{S^{i,z}}, f^{\lambda}_{S}) &= ||f^{\lambda}_{S^{i,z}}||^2_K - ||f^{\lambda}_{S}||^2_K - \langle f^{\lambda}_{S^{i,z}} - f^{\lambda}_{S}, \nabla ||f^{\lambda}_{S}||^2_K \rangle \\ &= ||f^{\lambda}_{S^{i,z}} - f^{\lambda}_{S}||^2_K \end{aligned}$$

Step 2

$$d_N(f^\lambda_{S^{i,z}}, f^\lambda_S) + d_N(f^\lambda_S, f^\lambda_{S^{i,z}}) \leq \frac{2L\|f^\lambda_S - f^\lambda_{S^{i,z}}\|_\infty}{\lambda n}$$

$$\begin{aligned}
\lambda(d_N(f^\lambda_{S^{i,z}}, f^\lambda_S) + d_N(f^\lambda_S, f^\lambda_{S^{i,z}})) &\leq \\
d_{T_S}(f^\lambda_{S^{i,z}}, f^\lambda_S) + d_{T_{S^{i,z}}}(f^\lambda_S, f^\lambda_{S^{i,z}}) &= \\
T_S(f^\lambda_{S^{i,z}}) - T_S(f^\lambda_S) + T_{S^{i,z}}(f^\lambda_S) - T_{S^{i,z}}(f^\lambda_{S^{i,z}}) &= \\
I_S(f^\lambda_{S^{i,z}}) - I_S(f^\lambda_S) + I_{S^{i,z}}(f^\lambda_S) - I_{S^{i,z}}(f^\lambda_{S^{i,z}}) &= \\
\frac{1}{n}(V(f^\lambda_{S^{i,z}}, z_i) - V(f^\lambda_S, z_i) + V(f^\lambda_S, z) - V(f^\lambda_{S^{i,z}}, z)) &\leq \\
\frac{2L\|f^\lambda_S - f^\lambda_{S^{i,z}}\|_\infty}{n}.
\end{aligned}$$

The Lemma

We want to prove that

$$||f^\lambda_{S^{i,z}} - f^\lambda_S||^2_K \leq \frac{2L\|f^\lambda_S - f^\lambda_{S^{i,z}}\|_\infty}{\lambda n}$$

The proof consists of two steps:
Step 1: prove that

$$2||f^\lambda_{S^{i,z}} - f^\lambda_S||^2_K = d_N(f^\lambda_{S^{i,z}}, f^\lambda_S) + d_N(f^\lambda_S, f^\lambda_{S^{i,z}})$$

Step 2: prove that

$$d_N(f^\lambda_{S^{i,z}}, f^\lambda_S) + d_N(f^\lambda_S, f^\lambda_{S^{i,z}}) \leq \frac{2L\|f^\lambda_S - f^\lambda_{S^{i,z}}\|_\infty}{\lambda n}$$

The Lemma

We want to prove that

$$||f^{\lambda}_{S^{i,z}} - f^{\lambda}_{S}||^2_K \leq \frac{2L\|f^{\lambda}_{S} - f^{\lambda}_{S^{i,z}}\|_{\infty}}{\lambda n}$$

The proof consists of two steps:
Step 1: prove that

$$2||f^{\lambda}_{S^{i,z}} - f^{\lambda}_{S}||^2_K = d_N(f^{\lambda}_{S^{i,z}}, f^{\lambda}_{S}) + d_N(f^{\lambda}_{S}, f^{\lambda}_{S^{i,z}})$$

Step 2: prove that

$$d_N(f^{\lambda}_{S^{i,z}}, f^{\lambda}_{S}) + d_N(f^{\lambda}_{S}, f^{\lambda}_{S^{i,z}}) \leq \frac{2L\|f^{\lambda}_{S} - f^{\lambda}_{S^{i,z}}\|_{\infty}}{\lambda n}$$

Stability of Tikhonov

1. assumption: $|V(f_1(x), y') - V(f_2(x), y')| \leq L\|f_1 - f_2\|_\infty$
2. property of RKHS: $\|f - f'\|_\infty \kappa \leq \|f - f'\|_K$, for any $f, f' \in \mathcal{H}$.
3. lemma: $||f_S^\lambda - f_{S^{i,z}}^\lambda||_K^2 \leq \frac{L\|f_S^\lambda - f_{S^{i,z}}^\lambda\|_\infty}{\lambda n}$

putting all together:

$$\begin{aligned} |V(f_S^\lambda, z) - V(f_{S^{z,i}}^\lambda, z)| &\leq L\|f_S^\lambda - f_{S^{z,i}}^\lambda\|_\infty \\ &\leq L\kappa||f_S^\lambda - f_{S^{z,i}}^\lambda||_K \\ &\leq \frac{L^2\kappa^2}{\lambda n} \\ &=: \beta \end{aligned}$$

Stability of Tikhonov

1. assumption: $|V(f_1(x), y') - V(f_2(x), y')| \leq L\|f_1 - f_2\|_\infty$
2. property of RKHS: $\|f - f'\|_\infty \kappa \leq \|f - f'\|_K$, for any $f, f' \in \mathcal{H}$.
3. lemma: $||f_S^\lambda - f_{S^{i,z}}^\lambda||_K^2 \leq \frac{L\|f_S^\lambda - f_{S^{i,z}}^\lambda\|_\infty}{\lambda n}$

putting all together:

$$
\begin{aligned}
|V(f_S^\lambda, z) - V(f_{S^{z,i}}^\lambda, z)| &\leq L\|f_S^\lambda - f_{S^{z,i}}^\lambda\|_\infty \\
&\leq L\kappa||f_S^\lambda - f_{S^{z,i}}^\lambda||_K \\
&\leq \frac{L^2\kappa^2}{\lambda n} \\
&=: \beta
\end{aligned}
$$

Stability of Tikhonov

1. assumption: $|V(f_1(x), y') - V(f_2(x), y')| \leq L\|f_1 - f_2\|_\infty$
2. property of RKHS: $\|f - f'\|_\infty \kappa \leq \|f - f'\|_K$, for any $f, f' \in \mathcal{H}$.
3. lemma: $||f_S^\lambda - f_{S^{i,z}}^\lambda||_K^2 \leq \frac{L\|f_S^\lambda - f_{S^{i,z}}^\lambda\|_\infty}{\lambda n}$

putting all together:

$$
\begin{aligned}
|V(f_S^\lambda, z) - V(f_{S^{z,i}}^\lambda, z)| &\leq L\|f_S^\lambda - f_{S^{z,i}}^\lambda\|_\infty \\
&\leq L\kappa||f_S^\lambda - f_{S^{z,i}}^\lambda||_K \\
&\leq \frac{L^2\kappa^2}{\lambda n} \\
&=: \beta
\end{aligned}
$$

Stability of Tikhonov

1. assumption: $|V(f_1(x), y') - V(f_2(x), y')| \leq L\|f_1 - f_2\|_\infty$
2. property of RKHS: $\|f - f'\|_\infty \kappa \leq \|f - f'\|_K$, for any $f, f' \in \mathcal{H}$.
3. lemma: $||f_S^\lambda - f_{S^{i,z}}^\lambda||_K^2 \leq \frac{L\|f_S^\lambda - f_{S^{i,z}}^\lambda\|_\infty}{\lambda n}$

putting all together:

$$
\begin{aligned}
|V(f_S^\lambda, z) - V(f_{S^{z,i}}^\lambda, z)| &\leq L\|f_S^\lambda - f_{S^{z,i}}^\lambda\|_\infty \\
&\leq L\kappa||f_S^\lambda - f_{S^{z,i}}^\lambda||_K \\
&\leq \frac{L^2\kappa^2}{\lambda n} \\
&=: \beta
\end{aligned}
$$

Bounding the Loss

We have shown that Tikhonov regularization with an L-Lipschitz loss is β-stable with $\beta = \frac{L^2\kappa^2}{\lambda n}$.
To apply the theorems and get the generalization bound, we need to bound the loss

$$V(f_S^\lambda, z) \leq M < \infty, \quad \forall z = (x, y)$$

We assume that

$$V(0, z) \leq C_0 < \infty$$

Bounding the Loss (cont.)

$$V(f_S^\lambda, z) \leq M < \infty, \quad \forall z = (x, y)$$

- Assume that $V(0, z) \leq C_0 < \infty$, then

$$\begin{aligned} \lambda ||f_S^\lambda||_K^2 &\leq T_S(f_S^\lambda) \leq T_S(0) \\ &= \frac{1}{n} \sum_{i=1}^{n} V(0, y_i) \leq C_0. \end{aligned}$$

- Then $||f_S^\lambda||_K^2 \leq \frac{C_0}{\lambda} \Longrightarrow \|f_S^\lambda\|_\infty \leq \kappa ||f_S^\lambda||_K \leq \kappa \sqrt{\frac{C_0}{\lambda}}$
- Finally $|V(f_S^\lambda(x), y)| \leq |V(f_S^\lambda(x), y) - V(0, y)| + |V(0, y)|$

$$|V(f_S^\lambda(x), y) - V(0, y)| \leq L \|f_S^\lambda\|_\infty \leq \kappa L \sqrt{\frac{C_0}{\lambda}}.$$

A Note on λ

We have shown that

$$\beta = \frac{L^2\kappa^2}{\lambda n}, \quad M = \kappa L\sqrt{\frac{C_0}{\lambda}} + C_0$$

so that, with probability $1 - \delta$,

$$I[f_S^\lambda] \leq I_S[f_S^\lambda] + \frac{L^2\kappa^2}{\lambda n} + (\frac{2L^2\kappa^2}{\lambda} + C_0 + \kappa L\sqrt{\frac{C_0}{\lambda}})\sqrt{\frac{2\ln(2/\delta)}{n}}.$$

Keeping λ fixed as n increase n, the generalization bound will tighten as $O\left(\frac{1}{\sqrt{n}}\right)$.
However, fixing λ fixed we keep our hypothesis space fixed. As we get more data, we want λ to get smaller. If λ gets smaller too fast, the bounds become trivial...