CS281B/Stat241B: Advanced Topics in Learning & Decision Making

Fourier Perspective on Regularization

Lecturer: Michael I. Jordan *Scribe: Amol Deshpande*

In this lecture, we will look at regularization from the Fourier perspective. In that case, the regularization expression is written in terms of a penalty in the frequency domain as follows (cf. Lecture : Regularization and RKHS) :

$$J_3 = c(x_i, y_i, f(x_i)) + \frac{\lambda}{2\pi} \int \frac{|\tilde{f}(\omega)|^2}{\nu(\omega)} d\omega$$

where $\tilde{f}(\omega) = \int f(x) e^{-i\omega x} dx$ is the Fourier transform of $f(x)$. We also have the dual property :

$$f(x) = \frac{1}{2\pi} \int \tilde{f}(\omega) e^{i\omega x} d\omega$$

where $\nu(\omega) = \nu(-\omega)$ is a symmetric real valued, non-negative function that converges to 0 as $|\omega| \to 0$. Smaller values of $\nu(\omega)$ pennalize the corresponding frequencies more and thus, $\nu(\omega)$ can be used to control the smoothness of the resulting $f(x)$.

Parseval's Theorem : Parseval's theorem relates the norm of a function $f(x)$ to the norm of the fourier transform of the function, $\tilde{f}(\omega)$:

$$\langle f, f \rangle = \int f(x) f^*(x) dx = \frac{1}{2\pi} \int |\tilde{f}(\omega)|^2 d\omega = \langle \tilde{f}, \tilde{f} \rangle$$

Proof :

$$\begin{aligned}
\langle f, f \rangle &= \int f(x) f^*(x) dx \\
&= \int \left\{ \frac{1}{2\pi} \int \tilde{f}(\omega) e^{i\omega x} d\omega \right\} \left\{ \frac{1}{2\pi} \int \tilde{f}(\omega') e^{i\omega' x} d\omega' \right\}^* dx \\
&= \left(\frac{1}{2\pi} \right)^2 \int d\omega d\omega' \tilde{f}(\omega) \tilde{f}^*(\omega') \int e^{ix(\omega - \omega')} dx \\
&= \left(\frac{1}{2\pi} \right)^2 \int d\omega d\omega' \tilde{f}(\omega) \tilde{f}^*(\omega') 2\pi \delta(\omega - \omega') \\
&= \frac{1}{2\pi} \int \tilde{f}^*(\omega) \tilde{f}(\omega) d\omega \\
&= \langle \tilde{f}, \tilde{f} \rangle
\end{aligned}$$

Fourier transforms for differential operators can be seen to be :

$$\begin{aligned}
\frac{d}{dx} f(x) &\quad \xrightarrow{FT} \quad i\omega \tilde{f}(\omega) \\
\frac{d^2}{dx^2} f(x) &\quad \xrightarrow{FT} \quad -\omega^2 \tilde{f}(\omega)
\end{aligned}$$

Using this, for the loss function $||\frac{d}{dx}f(x)||^2$, we have :

$$\int f'(x)^2 dx = \langle f', f' \rangle = \langle \tilde{f}', \tilde{f}' \rangle = \omega^2 \langle \tilde{f}, \tilde{f} \rangle = \frac{1}{2\pi} \int \omega^2 |\tilde{f}(\omega)|^2 d\omega$$

Similarly,

$$\int f''(x)^2 dx = \frac{1}{2\pi} \int \omega^4 |\tilde{f}(\omega)|^2 d\omega$$

In general, for a differential operator we get that :

$$||Lf||^2 = \frac{1}{2\pi} \int \frac{|\tilde{f}(\omega)|^2}{\nu(\omega)} d\omega$$

Thus, choosing L is equivalent to choosing $\nu(\omega)$.

To tie this back to RKHS, we use Green's functions. First we define a dot-product norm on this space :

$$\langle Lf, Lg \rangle = \frac{1}{2\pi} \int \frac{\tilde{f}(\omega)\overline{\tilde{g}(\omega)}}{\nu(\omega)} d\omega$$

We will make a guess as to what the Green's function might look like and show that we get the above norm using that.

Let $G(x, x_i) = G(x - x_i) = \frac{1}{2\pi} \int e^{i\omega(x - x_i)} \nu(\omega) d\omega$. Then the Fourier transform of G is :

$$\begin{aligned}
\tilde{G}(\omega) &= \frac{1}{2\pi} \int dx e^{-i\omega x} \int d\omega' e^{i\omega'(x - x_i)} \nu(\omega') \\
&= \frac{1}{2\pi} \int d\omega' e^{-i\omega' x_i} \nu(\omega') \int dx e^{-i(\omega - \omega')x} \\
&= \frac{1}{2\pi} \int d\omega' e^{-i\omega' x_i} \nu(\omega') 2\pi \delta(\omega - \omega') \\
&= e^{-i\omega x_i} \nu(\omega)
\end{aligned}$$

Finally to see whether we have a RKHS :

$$\begin{aligned}
\langle LG(x, x_i), LG(x, x_j) \rangle &= \frac{1}{2\pi} \int \frac{e^{-i\omega x_i} \nu(\omega) e^{i\omega x_j} \nu(\omega)}{\nu(\omega)} d\omega \\
&= \frac{1}{2\pi} \int e^{-i\omega(x_i - x_j)} \nu(\omega) d\omega \\
&= G(x_i, x_j)
\end{aligned}$$

Thus, G satisfies the reproducing kernel property.

Thus, the regularization term in the loss function can be thought of as the norm of a function in an RKHS defined by the kernel $G(x - x_i)$. Finally, we note that $f(x) = \Sigma_i \alpha_i G(x - x_i)$ (by the Representer Theorem).