MATH 690: Topics in Data Analysis and Computation November 28, 2017

Stability of convolutional neural network

Lecturer: Xiuyuan Cheng *Scriber: Zuoming Dai, Tianyu Wang*

1 Stability

Recall that, in the simplest case, a one-layer convolutional neural network can be written as the following:

$$y(u) = \sigma(x * \psi(u)) \tag{1}$$

where x is the signal, ψ is the filter and σ is a transformation that is applied at every neuron. Usually, σ is sigmoid function or ReLU.

1.1 L_2 stability

The L_2 stability comes in a Lipchitz format. More specifically, a function y is L_2 stable if

$$\|y(x) - y(\tilde{x})\|_2 \leq C \|x - \tilde{x}\|_2$$

for any choice of x and $\tilde{x}$. Next, we are going to show that (1) is L_2 stable under certain conditions, and thus by induction, a multi-layer CNN is also L_2 stable.

First, when σ is non-expansive, i.e., $|\sigma(z) - \sigma(z')| \leq |z - z'|$ for any z and z',

$$\begin{aligned}\|y(x) - y(\tilde{x})\|_2^2 &= \int |\sigma(x * \psi(u)) - \sigma(\tilde{x} * \psi(u))|^2 \, du \\ &\leq \int |x * \psi(u) - \tilde{x} * \psi(u)|^2 \, du \qquad (2)\end{aligned}$$

Next, we will show that (2) is L_2 stable under the mild assumption that $\|\psi\|_1$ is bounded.

Proposition 1.

$$\|x * \psi\|_2 \leq \|\psi\|_1 \|x\|_2$$

Proof.

$$\begin{aligned}\|x * \psi\|_2^2 &= \int |x * \psi(u)|^2 \, du \\ &= \int \left| \int x(v)\psi(u - v) dv \right|^2 du \\ &\leq \int \left(\int |x(v)|^2 |\psi(u - v)| \, dv \right) \left(\int |\psi(u - v)| \, dv \right) du \qquad (3) \\ &= \|x\|_2^2 \|\psi\|_1^2\end{aligned}$$

■

Here (3) is a consequence of the Cauchy-Schwarz inequality on $x(v)\psi(u - v)^{\frac{1}{2}}$ and $\psi(u - v)^{\frac{1}{2}}$. Note that we do not apply the Cauchy-Schwarz inequality to $x(v)$ and $\psi(u - v)$ directly, since the Hilbert-Schmidt condition is often not preferred. Recall that a integration operator is defined as the following.

Definition 1 (Integration Operator)**.** *For a real-valued kernel function $k(\cdot,\cdot)$ defined over measurable space $\mathcal{X} \times \mathcal{Y}$, the integration operator defined with respect to k is defined as*

$$T_k f(u) = \int_Y k(u,v) f(v) dv$$

In addition, a kernel $k(\cdot,\cdot)$ is said to be Hilbert-Schmidt if $\int \int |k(u,v)|^2 \, dvdu < \infty$.

It is easy to check that the (translation-invariant) hat function $k(u,v) = \mathbf{1}_{[-1,1]}(u-v)$ is not Hilbert-Schmidt. On the other hand, its 1-norm is finite. Therefore, applying the Cauchy-Schwarz inequality to $x(v)$ and $\psi(u-v)$ directly will result in a bound that is looser than applying the Cauchy-Schwarz inequality to $x(v)\psi(u-v)^{\frac{1}{2}}$ and $\psi(u-v)^{\frac{1}{2}}$.
In fact, Proposition 1 is a special case of Theorem 1.

Theorem 1 (Schur's Test)**.** *Suppose T_k is the integration operator with kernel $k(\cdot,\cdot)$. If there exists $C > 0$ s.t.*

$$\begin{cases} \int |k(u,v)| dv \leq C, & \forall \, u \\ \int |k(u,v)| du \leq C, & \forall \, v \end{cases}$$

then $\|T_k\|_2 \leq C$, i.e., $\|T_k f\|_2 \leq C \|f\|_2$. Here $\|T_k\|_2 = \sup_{f, \|f\|_2 = 1} \|T_k f\|_2$.