Neural Network Learning: Theoretical Foundations

Speaker : Semin Choi

Department of Statistics, Seoul National University, South Korea

November 22, 2017

Comparing the Different Approaches

Bounding with the Fat-Shattering Dimension

- Theorem 12.8.

$$\mathcal{N}_1(\epsilon, F, m) \leq \mathcal{N}_\infty(\epsilon, F, m) \leq \left(\frac{\sqrt{m}}{\epsilon}\right)^{\mathrm{fat}_F(\epsilon/4)\log_2(m/(\epsilon \mathrm{fat}_F(\epsilon/4)))}.$$

- Theorem 18.2.

$$\mathcal{N}_1(\epsilon, F, m) \leq \left(\frac{1}{\epsilon}\right)^{\mathrm{fat}_F(\epsilon/8)\log_2(m/(\epsilon \mathrm{fat}_F(\epsilon/8)))}$$

Bounding with the Pseudo-Dimension

- Theorem 12.2.

$$\mathcal{N}_1(\epsilon, F, m) \leq \mathcal{N}_\infty(\epsilon, F, m) \leq \left(\frac{m}{\epsilon}\right)^{\mathrm{Pdim}(F)}$$

- Theorem 18.4.

$$\mathcal{N}_1(\epsilon, F, m) \leq \left(\frac{1}{\epsilon}\right)^{\mathrm{Pdim}(F)}$$

Uniform Convergence Results

Part 2 (Classification with Real-valued Functions)

- $\mathrm{er}_P(f) = P\{\mathrm{sgn}(f(x) - 1/2) \neq y\}$
- $\mathrm{er}_P^\gamma(f) = P\{\mathrm{margin}(f(x), y) < \gamma\}$

$$P^m\{\mathrm{er}_P(f) \geq \hat{\mathrm{er}}_z^\gamma(f) + \epsilon \text{ for some } f \text{ in } F\} \leq 2\mathcal{N}_\infty(\gamma/2, F, 2m)\exp\left(-\frac{\epsilon^2 m}{8}\right).$$

Part 3 (Regression)

- $\mathrm{er}_P(f) = \mathbb{E}(f(x) - y)^2 = \mathbb{E}l_f$

$$P^m\{|\mathrm{er}_P(f) - \hat{\mathrm{er}}_z(f)| \geq \epsilon \text{ for some } f \text{ in } F\} \leq 4\mathcal{N}_1(\epsilon/16, F, 2m)\exp\left(-\frac{\epsilon^2 m}{32}\right).$$

Proof

Part 2

Lemma 1.

$$\max_{z \in Z^{2m}} \Pr(\sigma z \in R) \leq \mathcal{N}_\infty(\gamma/2, F, 2m) \exp\left(-\frac{\epsilon^2 m}{8} \right)$$

where

$$R = \{(r, s) \in Z^m \times Z^m : \text{ some } f \text{ in } F \text{ has } \hat{\text{er}}_s(f) \geq \hat{\text{er}}_r^\gamma(f) + \epsilon/2\}$$

If $\sigma z \in R$, then there is some $\hat{f} \in T$ such that

$$\hat{\text{er}}_s^{\gamma/2}(\hat{f}) \geq \hat{\text{er}}_r^{\gamma/2}(\hat{f}) + \epsilon/2.$$

Proof

Part 3

Lemma 2.

$$\max_{z \in Z^{2m}} \Pr(\sigma z \in R) \leq 2\mathcal{N}_1(\epsilon/16, F, 2m) \exp\left(-\frac{\epsilon^2 m}{32}\right)$$

where

$$R = \{(r, s) \in Z^m \times Z^m : \text{ some } f \text{ in } F \text{ has } |\hat{\text{er}}_r(f) - \hat{\text{er}}_s(f)| \geq \epsilon/2\}$$

If $\sigma z \in R$, then there is some $\hat{f} \in T$ such that

$$\left|\hat{\text{er}}_r(\hat{f}) - \hat{\text{er}}_s(\hat{f})\right| \geq \epsilon/4.$$