PAC-Bayes-Empirical-Bernstein Inequality

Ilya Tolstikhin[1] Yevgeny Seldin[2]

[1]Computing Centre of Russian Academy of Sciences

[2]Queensland University of Technology
UC Berkeley

NIPS
Decemper 2013

Poster ID: Sat39

Basic Definitions

$\mathcal{X}$ — sample space

$\mathcal{Y}$ — label space

$\ell : \mathcal{Y} \times \mathcal{Y} \rightarrow [0, 1]$ — loss function

$\mathcal{H}$ — hypothesis space

$h(x)$ — prediction of h on x

$L(h) = \mathbb{E}_{(x,y)\sim\mathcal{D}}[\ell(y, h(x))]$

$L_n(h) = \frac{1}{n}\sum_{i=1}^{n} \ell(y_i, h(x_i))$

$\mathbb{V}(h) = \mathrm{Var}[\ell(y, h(x))]$

$\mathbb{V}_n(h) = \frac{\sum_{i,j}(\ell(y_i,h(x_i)) - \ell(y_j,h(x_j)))^2}{2n(n-1)}$

Randomized Prediction Rule

ρ — distribution over $\mathcal{H}$.
At each round of the game:

1. Pick $h \sim \rho$
2. Observe x
3. Return $h(x)$

Example: ρ - Bayes posterior

Loss of ρ

$$L(\rho) = \mathbb{E}_{h\sim\rho}[L(h)]$$
$$L_n(\rho) = \mathbb{E}_{h\sim\rho}[L_n(h)]$$
$$\mathbb{V}(\rho) = \mathbb{E}_{h\sim\rho}[\mathbb{V}(h)]$$
$$\mathbb{V}_n(\rho) = \mathbb{E}_{h\sim\rho}[\mathbb{V}_n(h)]$$

The Tightest PAC-Bayesian Bounds

For any fixed π with high probability for all ρ simultaneously:

PAC-Bayes-kl (Seeger, 2002)

$$\mathrm{kl}(L_n(\rho)\|L(\rho)) \leq \frac{\mathrm{KL}(\rho\|\pi) + o(..)}{n}$$

Relaxed PAC-Bayes-kl

$$L(\rho) \leq L_n(\rho) + \sqrt{\frac{2L_n(\rho)\left(\mathrm{KL}(\rho\|\pi) + o(..)\right)}{n}} + \frac{2\left(\mathrm{KL}(\rho\|\pi) + o(..)\right)}{n}$$

PAC-Bayes-Bernstein (Seldin et. al., 2012)

$$L(\rho) \leq L_n(\rho) + \sqrt{\frac{3\mathbb{V}(\rho)\left(\mathrm{KL}(\rho\|\pi) + o(..)\right)}{n}}$$

- Problem: $\mathbb{V}(\rho)$ is unknown
- $L_n(\rho)$ is a potentially lose upper bound on $\mathbb{V}(\rho)$

Our results

Poster ID: Sat39

- **PAC-Bayesian bound on $\mathbb{V}(\rho)$:**

$$\mathbb{V}(\rho) \leq \underbrace{\mathbb{V}_n(\rho) + \sqrt{\frac{2\mathbb{V}_n(\rho)\left(\mathrm{KL}(\rho\|\pi) + o(..)\right)}{n-1}} + \frac{2\left(\mathrm{KL}(\rho\|\pi) + o(..)\right)}{n-1}}_{\bar{V}(\rho)}$$

- **PAC-Bayes-Empirical-Bernstein inequality:**

$$L(\rho) \leq L_n(\rho) + \sqrt{\frac{3\bar{V}(\rho)\left(\mathrm{KL}(\rho\|\pi) + o(..)\right)}{n}}$$

PB-EB / PB-kl