1 PAC Learning

Empirical error: $\hat{\mathcal{R}}_n(c) = \frac{1}{n}\sum_{i=1}^n \mathbb{I}_{\{c(x_i)\neq y\}}$

Expected error: $\mathcal{R}(c) = P\{c(x) \neq y\}$

ERM: $\hat{c}_n^* = \arg\min_{c\in\mathcal{C}} \hat{\mathcal{R}}_n(c)$

opt: $c^* \in \min_{c\in\mathcal{C}} \mathcal{R}(c)$, $|\mathcal{C}|$ finite

Generalization error: $\mathcal{R}(\hat{c}_n^*) = P\{\hat{c}_n^*(x) \neq y\}$

VC ineq.: $\mathcal{R}(\hat{c}_n^*) - \inf_{c\in\mathcal{C}} \mathcal{R}(c) \leq 2\sup_{c\in\mathcal{C}} |\hat{\mathcal{R}}_n(c) - \mathcal{R}(c)|$

$P\{\mathcal{R}(\hat{c}_n^*) - \mathcal{R}(c^*) > \epsilon\} \leq P\{\sup_{c\in\mathcal{C}} |\hat{\mathcal{R}}_n(c)-\mathcal{R}(c)| > \frac{\epsilon}{2}\}$

$\leq 2|\mathcal{C}| exp(-2n\epsilon^2/4) \leq 8s(\mathcal{A},n)exp(-n\epsilon^2/32)$

and $s(\mathcal{A},n) \leq n^{\mathcal{V}_\mathcal{A}}$

Markov ineq: $P\{X\geq\epsilon\} \leq \frac{\mathbb{E}[X]}{\epsilon}$ (for nonneg. X)

Boole's inequality: $P(\bigcup_i A_i) \leq \sum_i P(A_i)$

Hoeffding's lemma: $\mathbb{E}[e^{sX}] \leq exp(\frac{1}{8}s^2(b-a)^2)$

where $\mathbb{E}[X]=0$, $P(X\in[a,b])=1$

Hoeffding's: $P\{S_n - \mathbb{E}[S_n] \geq t\} \leq exp(-\frac{2t^2}{\sum_i (b_i-a_i)^2})$

Normalized: $P\{\widetilde{S}_n - \mathbb{E}[\widetilde{S}_n] \geq \epsilon\} \leq exp(-\frac{2n^2\epsilon^2}{\sum_i(b_i-a_i)^2})$

Error bound: $P\{\sup_{c\in\mathcal{C}}|\hat{\mathcal{R}}_n(c) - \mathcal{R}(c)| > \epsilon\} \leq 2|\mathcal{C}|exp(-2n\epsilon^2)$

The $\mathcal{VC}$ dimension of a model f is the maximum number of points that can be arranged so that f shatters them.

2 PAC Learning

Empirical error: $\hat{\mathcal{R}}_n(c) = \frac{1}{n}\sum_{i=1}^n \mathbb{I}_{\{c(x_i)\neq y\}}$

Expected error: $\mathcal{R}(c) = P\{c(x) \neq y\}$

ERM: $\hat{c}_n^* = \arg\min_{c\in\mathcal{C}} \hat{\mathcal{R}}_n(c)$

opt: $c^* \in \min_{c\in\mathcal{C}} \mathcal{R}(c)$, $|\mathcal{C}|$ finite

Generalization error: $\mathcal{R}(\hat{c}_n^*) = P\{\hat{c}_n^*(x) \neq y\}$

VC ineq.: $\mathcal{R}(\hat{c}_n^*) - \inf_{c\in\mathcal{C}} \mathcal{R}(c) \leq 2\sup_{c\in\mathcal{C}} |\hat{\mathcal{R}}_n(c) - \mathcal{R}(c)|$

$P\{\mathcal{R}(\hat{c}_n^*) - \mathcal{R}(c^*) > \epsilon\} \leq P\{\sup_{c\in\mathcal{C}} |\hat{\mathcal{R}}_n(c)-\mathcal{R}(c)| > \frac{\epsilon}{2}\}$

$\leq 2|\mathcal{C}| exp(-2n\epsilon^2/4) \leq 8s(\mathcal{A},n)exp(-n\epsilon^2/32)$

and $s(\mathcal{A},n) \leq n^{\mathcal{V}_\mathcal{A}}$

Markov ineq: $P\{X\geq\epsilon\} \leq \frac{\mathbb{E}[X]}{\epsilon}$ (for nonneg. X)

Boole's inequality: $P(\bigcup_i A_i) \leq \sum_i P(A_i)$

Hoeffding's lemma: $\mathbb{E}[e^{sX}] \leq exp(\frac{1}{8}s^2(b-a)^2)$

where $\mathbb{E}[X]=0$, $P(X\in[a,b])=1$

Hoeffding's: $P\{S_n - \mathbb{E}[S_n] \geq t\} \leq exp(-\frac{2t^2}{\sum_i (b_i-a_i)^2})$

Normalized: $P\{\widetilde{S}_n - \mathbb{E}[\widetilde{S}_n] \geq \epsilon\} \leq exp(-\frac{2n^2\epsilon^2}{\sum_i(b_i-a_i)^2})$

Error bound: $P\{\sup_{c\in\mathcal{C}}|\hat{\mathcal{R}}_n(c) - \mathcal{R}(c)| > \epsilon\} \leq 2|\mathcal{C}|exp(-2n\epsilon^2)$

The $\mathcal{VC}$ dimension of a model f is the maximum number of points that can be arranged so that f shatters them.

3 PAC Learning

Empirical error: $\hat{\mathcal{R}}_n(c) = \frac{1}{n}\sum_{i=1}^n \mathbb{I}_{\{c(x_i)\neq y\}}$

Expected error: $\mathcal{R}(c) = P\{c(x) \neq y\}$

ERM: $\hat{c}_n^* = \arg\min_{c\in\mathcal{C}} \hat{\mathcal{R}}_n(c)$

opt: $c^* \in \min_{c\in\mathcal{C}} \mathcal{R}(c)$, $|\mathcal{C}|$ finite

Generalization error: $\mathcal{R}(\hat{c}_n^*) = P\{\hat{c}_n^*(x) \neq y\}$

VC ineq.: $\mathcal{R}(\hat{c}_n^*) - \inf_{c\in\mathcal{C}} \mathcal{R}(c) \leq 2\sup_{c\in\mathcal{C}} |\hat{\mathcal{R}}_n(c) - \mathcal{R}(c)|$

$P\{\mathcal{R}(\hat{c}_n^*) - \mathcal{R}(c^*) > \epsilon\} \leq P\{\sup_{c\in\mathcal{C}} |\hat{\mathcal{R}}_n(c)-\mathcal{R}(c)| > \frac{\epsilon}{2}\}$

$\leq 2|\mathcal{C}| exp(-2n\epsilon^2/4) \leq 8s(\mathcal{A},n)exp(-n\epsilon^2/32)$

and $s(\mathcal{A},n) \leq n^{\mathcal{V}_\mathcal{A}}$

Markov ineq: $P\{X\geq\epsilon\} \leq \frac{\mathbb{E}[X]}{\epsilon}$ (for nonneg. X)

Boole's inequality: $P(\bigcup_i A_i) \leq \sum_i P(A_i)$

Hoeffding's lemma: $\mathbb{E}[e^{sX}] \leq exp(\frac{1}{8}s^2(b-a)^2)$

where $\mathbb{E}[X]=0$, $P(X\in[a,b])=1$

Hoeffding's: $P\{S_n - \mathbb{E}[S_n] \geq t\} \leq exp(-\frac{2t^2}{\sum_i (b_i-a_i)^2})$

Normalized: $P\{\widetilde{S}_n - \mathbb{E}[\widetilde{S}_n] \geq \epsilon\} \leq exp(-\frac{2n^2\epsilon^2}{\sum_i(b_i-a_i)^2})$

Error bound: $P\{\sup_{c\in\mathcal{C}}|\hat{\mathcal{R}}_n(c) - \mathcal{R}(c)| > \epsilon\} \leq 2|\mathcal{C}|exp(-2n\epsilon^2)$

The $\mathcal{VC}$ dimension of a model f is the maximum number of points that can be arranged so that f shatters them.