MATH466/MATH766
Math of machine learning
Appendix VC Inequality and Error of Binary Classification

References:

•

Todays contents:

- VC inequality
- A corollary

Important concepts:

•

Recommend reading:

•

Binary Classification

$\mathcal{H}$ is a collection of binary classifiers on $\mathcal{X}$

$$\mathcal{H} := \{ h : \mathcal{X} \to \{0,1\} \}$$

Risk : $R(h) = \mathbb{E}_{(X,Y) \sim P_{X,Y}} \mathbb{1}_{h(X) \neq Y}$

Emp Risk : $\hat{R}_n(h) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{h(X_i) \neq Y}$, $(X_i, Y_i) \sim_{iid} P_{X,Y}$

Let $\hat{h}_n := \underset{h \in \mathcal{H}}{\operatorname{argmin}} \hat{R}_n(h)$ (empirical minimizer)

want to know $\underbrace{R(\hat{h}_n)}_{\text{random}} - \min_{h \in \mathcal{H}} R(h)$

1. Want to know $\Pr[R(\hat{h}_n) \leq \min_{h \in \mathcal{H}} R(h) + \varepsilon]$

Denote $h^* := \underset{h \in \mathcal{H}}{\operatorname{argmin}} R(h)$

$R(\hat{h}_n) - \min_{h \in \mathcal{H}} R(h)$

$= \left(R(\hat{h}_n) - \hat{R}_n(\hat{h}_n) \right) + \underbrace{\left(\hat{R}_n(\hat{h}_n) - \hat{R}_n(h^*) \right)}_{\leq 0} + \left(\hat{R}_n(h^*) - R(h^*) \right)$

$\leq | R(\hat{h}_n) - \hat{R}_n(\hat{h}_n) | + | \hat{R}_n(h^*) - R(h^*) |$

$\leq 2 \sup_{h \in \mathcal{H}} | \hat{R}_n(h) - R(h) |$

2. Want to know $\Pr\left[\sup_{h\in\mathcal{H}} |\hat{R}_n(h) - R(h)| \le \varepsilon\right]$

Denote $\mathcal{A} := \{A_h : h \in \mathcal{H}\}$

where $A_h := \{(x,y) \in \mathcal{X} \times \{0,1\} : h(x) \neq y\}$

the set of misclassification cases under h

then $R(h) = \Pr[(X,Y) \in A_h \mid (X,Y) \sim P_{X,Y}] =: \mu(A_h)$

$\hat{R}_n(h) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{(X_i,Y_i)\in A_h} =: \mu_n(A_h)$

Shown in lec 20 : $V_{\mathcal{A}} = V_{\mathcal{H}}$

3. Assume $Z \sim P_Z$, $Z_1, Z_2, \dots, Z_n \sim_{iid} P_Z$

For a set A, denote $\mu(A) := \Pr[Z \in A]$

$\mu_n(A) := \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{Z_i\in A}$

Want to know $\Pr\Big[\underbrace{\sup_{A\in\mathcal{A}} |\mu_n(A) - \mu(A)|}_{\text{r.v.}} \le \varepsilon\Big]$

4. Bounded Difference Inequality (McDiarmid's inequality)

Def (bounded difference property) $f : \mathcal{Z}_1 \times \mathcal{Z}_2 \times \cdots \times \mathcal{Z}_n \to \mathbb{R}$

$\exists\, c_1, c_2, \dots, c_n$ s.t.

for all $i = 1, 2, \dots, n$ and all $z_1 \in \mathcal{Z}_1, z_2 \in \mathcal{Z}_2, \dots, z_n \in \mathcal{Z}_n$

$$\sup_{z_i' \in \mathcal{Z}_i} \left| f(z_1, z_2, \dots, z_i', \dots, z_n) - f(z_1, z_2, \dots, z_i, \dots, z_n)\right| \le c_i$$

(substituting the i-th coordinate changes the value of f at most c_i)

Thm. Let f satisfies BDP w/. $c_1, c_2, \dots, c_n$

$Z_1, Z_2, \dots, Z_n$ are independent r.v., $Z_i \in \mathcal{Z}_i$ for all i

Then for any $\varepsilon > 0$

$$\Pr[f(Z_1, \dots, Z_n) - \mathbb{E}[f(Z_1, \dots, Z_n)] \geq \varepsilon]$$
$$=$$
$$\Pr[f(Z_1, \dots, Z_n) - \mathbb{E}[f(Z_1, \dots, Z_n)] \leq -\varepsilon] \leq \exp\left(-\frac{2\varepsilon^2}{\sum_{i=1}^n c_i^2}\right)$$

In our setting. $c_1 = c_2 = \dots = c_n = \frac{1}{n}$

5. VC inequality

Let $d := V_{\mathcal{A}}$

then $$\mathbb{E}\left[\sup_{A \in \mathcal{A}} |\mu_n(A) - \mu(A)|\right] \leq 2\sqrt{\frac{2d \log(2en/d)}{n}}$$

$$\left.\begin{matrix}\boxed{4} \\ \boxed{5}\end{matrix}\right] - \boxed{3} - \boxed{2} - \boxed{1}$$

With probability $1-\delta$

$$R(\hat{h}_n) \leq \min_{h \in \mathcal{H}} R(h) + 4\sqrt{\frac{2d \log(2en/d)}{n}} + \sqrt{\frac{\log(2/\delta)}{2n}}$$