12.4

$$f(E(x)) \leq E(f(x))$$

[推导]:显然,对于任意凸函数,必然有:

$$f\left(\alpha x_1 + (1 - \alpha)x_2\right) \le \alpha f\left(x_1\right) + (1 - \alpha)f\left(x_2\right)$$

$$f(E(x)) = f\left(rac{1}{m}\sum_i^m x_i
ight) = f\left(rac{m-1}{m}rac{1}{m-1}\sum_i^{m-1} x_i + rac{1}{m}x_i
ight)$$

取:

$$\alpha = \frac{m-1}{m}$$

所以推得:

$$f(E(x)) \leq rac{m-1}{m} f\left(rac{1}{m-1} \sum_{i}^{m-1} x_i
ight) + rac{1}{m} f\left(x_m
ight)$$

以此类推得:

$$f(E(x)) \leq rac{1}{m}f\left(x_1
ight) + rac{1}{m}f\left(x_2
ight) + \ldots \ldots + rac{1}{m}f\left(x_m
ight) = E(f(x))$$

12.17

$$P(|\hat{E}(h) - E(h)| \ge \varepsilon) \le 2e^{-2m\varepsilon^2}$$

[推导]: 已知Hoeffding不等式:若 $x_1,x_2\ldots x_m$ 为m个独立变量,且满足 $0\leq x_i\leq 1$,则对任意 $\varepsilon>0$,有:

$$P\left(\left|rac{1}{m}\sum_{i}^{m}x_{i}-rac{1}{m}\sum_{i}^{m}E\left(x_{i}
ight)
ight|\geqarepsilon
ight)\leq2e^{-2marepsilon^{2}}$$

将 x_i 替换成损失函数 $l\left(h\left(x_i\right) \neq y_i\right)$,显然 $0 \leq l\left(h\left(x_i\right) \neq y_i\right) \leq 1$,且独立,带入Hoeffiding不等式可得:

$$P\left(|rac{1}{m}\sum_{i}^{m}l\left(h\left(x_{i}
ight)
eq y_{i}
ight)-rac{1}{m}\sum_{i}^{m}E\left(l\left(x_{i}
ight)
eq y_{i}
ight)
ight)|\geqarepsilon
ight)\leqarepsilon^{-2marepsilon^{2}}$$

其中:

$$\hat{E}(h) = rac{1}{m} \sum_{i}^{m} l\left(h\left(x_{i}
ight)
eq y_{i}
ight)$$

$$E(h) = P_{x \in \mathbb{D}} l(h(x)
eq y) = E(l(h(x)
eq y)) = rac{1}{m} \sum_{i}^{m} E\left(l\left(h\left(x_{i}
ight)
eq y_{i}
ight)
ight)$$

所以有:

$$P(|\hat{E}(h) - E(h)| \ge arepsilon) \le 2e^{-2marepsilon^2}$$

12.18

$$\hat{E}(h) - \sqrt{rac{\ln(2/\delta)}{2m}} \leq E(h) \leq \hat{E}(h) + \sqrt{rac{\ln(2/\delta)}{2m}}$$

[推导]:由(12.17)可知:

$$P(|\hat{E}(h) - E(h)| \ge \varepsilon) \le 2e^{-2m\varepsilon^2}$$

成立

即:

$$P(|\hat{E}(h) - E(h)| \le \varepsilon) \ge 1 - 2e^{-2m\varepsilon^2}$$

取
$$\delta=2e^{-2marepsilon^2}$$
,则 $arepsilon=\sqrt{rac{\ln(2/\delta)}{2m}}$

所以
$$|\hat{E}(h) - E(h)| \leq \sqrt{rac{\ln(2/\delta)}{2m}}$$
的概率不小于 $1 - \delta$

整理得:

$$\hat{E}(h) - \sqrt{rac{\ln(2/\delta)}{2m}} \leq E(h) \leq \hat{E}(h) + \sqrt{rac{\ln(2/\delta)}{2m}}$$

以至少 $1 - \delta$ 的概率成立

12.59

$$l(arepsilon,D) \leq l_{loo}(\overline{arepsilon},D) + eta + (4meta + M)\sqrt{rac{\ln(1/\delta)}{2m}}$$

[解析]:取 $\varepsilon=\beta+(4m\beta+M)\sqrt{\frac{\ln(1/\delta)}{2m}}$ 时,可以得到:

 $l(\varepsilon,D)-l_{loo}(\varepsilon,D)\leq \varepsilon$ 以至少 $1-\frac{\delta}{2}$ 的概率成立,K折交叉验证,当K=m时,就成了留一法,这时候会有很不错的泛化能力,但是有前提条件,对于损失函数l满足 β 均匀稳定性,且 β 应该是 $O(\frac{1}{m})$ 这个量级,仅拿出一个样本,可以保证很小的 β ,随着K的减小,训练的样本会减少, β 会逐渐增大,当 β 量级小于 $O(\frac{1}{m})$ 时,交叉验证就会不合理了

附录

给定函数空间 F_1, F_2 ,证明Rademacher复杂度:

$$R_m\left(F_1+F_2
ight) \leq R_m\left(F_1
ight) + R_m\left(F_2
ight)$$

[推导]:

$$R_{m}\left(F_{1}+F_{2}
ight)=E_{Z\in\mathbf{z}:\left|Z
ight|=m}\left[\hat{R}_{Z}\left(F_{1}+F_{2}
ight)
ight]$$

$$\hat{R}_{Z}\left(F_{1}+F_{2}
ight)=E_{\sigma}\left[\sup_{f_{1}F_{1},f_{2}\in F_{2}}rac{1}{m}\sum_{i}^{m}\sigma_{i}\left(f_{1}\left(z_{i}
ight)+f_{2}\left(z_{i}
ight)
ight)
ight]$$

当 $f_1(z_i) f_2(z_i) < 0$ 时,

$$\sigma_{i}\left(f_{1}\left(z_{i}
ight)+f_{2}\left(z_{i}
ight)
ight)<\sigma_{i1}f_{1}\left(z_{i}
ight)+\sigma_{i2}f_{2}\left(z_{i}
ight)$$

当 $f_{1}\left(z_{i}\right)f_{2}\left(z_{i}\right)\geq0$ 时,

$$\sigma_{i}\left(f_{1}\left(z_{i}
ight)+f_{2}\left(z_{i}
ight)
ight)=\sigma_{i1}f_{1}\left(z_{i}
ight)+\sigma_{i2}f_{2}\left(z_{i}
ight)$$

所以:

$$\hat{R}_{Z}\left(F_{1}+F_{2}
ight)\leq\hat{R}_{Z}\left(F_{1}
ight)+\hat{R}_{Z}\left(F_{2}
ight)$$

也即:

$$R_m\left(F_1+F_2\right)\leq R_m\left(F_1\right)+R_m\left(F_2\right)$$