## 12.4

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

[推导]：显然，对于任意凸函数，必然有：

$$f\left(\alpha x_{1}+(1-\alpha) x_{2}\right) \leq \alpha f\left(x_{1}\right)+(1-\alpha) f\left(x_{2}\right)$$

$$f(E(x))=f\left(\frac{1}{m} \sum_{i}^{m} x_{i}\right)=f\left(\frac{m-1}{m} \frac{1}{m-1} \sum_{i}^{m-1} x_{i}+\frac{1}{m} x_{i}\right)$$

取：

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

所以推得：

$$f(E(x)) \leq \frac{m-1}{m} f\left(\frac{1}{m-1} \sum_{i}^{m-1} x_{i}\right)+\frac{1}{m} f\left(x_{m}\right)$$

以此类推得：

$$f(E(x)) \leq \frac{1}{m} f\left(x_{1}\right)+\frac{1}{m} f\left(x_{2}\right)+\ldots \ldots+\frac{1}{m} f\left(x_{m}\right)=E(f(x))$$

## 12.17

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

[推导]：已知Hoeffding不等式：若$x_{1}, x_{2} \ldots \ldots x_{m}$为$m$个独立变量，且满足$0 \leq x_{i} \leq 1$ ,则对任意$\varepsilon>0$,有：

$$P\left(\left|\frac{1}{m} \sum_{i}^{m} x_{i}-\frac{1}{m} \sum_{i}^{m} E\left(x_{i}\right)\right| \geq \varepsilon\right) \leq 2 e^{-2 m \varepsilon^{2}}$$

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Hoeffding不等式可得：

$$P\left(| \frac{1}{m} \sum_{i}^{m} l\left(h\left(x_{i}\right) \neq y_{i}\right)-\frac{1}{m} \sum_{i}^{m} E\left(l\left(x_{i}\right) \neq y_{i}\right)\right) | \geq \varepsilon) \leq 2 e^{-2 m \varepsilon^{2}}$$

其中：

$$\hat{E}(h)=\frac{1}{m} \sum_{i}^{m} l\left(h\left(x_{i}\right) \neq y_{i}\right)$$

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

所以有：

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

## 12.18

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

[推导]：由（12.17）可知：

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

成立

即:

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

取$\delta=2e^{-2m\varepsilon^2}$,则$\varepsilon=\sqrt{\frac{\ln(2/\delta)}{2m}}$

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

整理得：

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

以至少$1-\delta$的概率成立

## 12.59

$$l(\varepsilon, D) \leq l_{loo}(\overline{\varepsilon}, D)+\beta+(4m\beta+M)\sqrt{\frac{\ln(1/\delta)}{2m}}$$

[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$\beta$均匀稳定性，且$\beta$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$\beta$,随着$K$的减小，训练的样本会减少，$\beta$会逐渐增大，当$\beta$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:

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

[推导]：

$$R_m\left(F_1+F_2\right)=E_{Z \in \mathbf{z}:|Z|=m}\left[\hat{R}_Z\left(F_1+F_2\right)\right]$$

$$\hat{R}_Z\left(F_1+F_2\right)=E_\sigma\left[\sup_{f_1 F_1, f_2 \in F_2} \frac{1}{m} \sum_i^m \sigma_i\left(f_1\left(z_i\right)+f_2\left(z_i\right)\right)\right]$$

当$f_1\left(z_i\right) f_2\left(z_i\right)<0$时，

$$\sigma_i\left(f_1\left(z_i\right)+f_2\left(z_i\right)\right)<\sigma_{i1}f_1\left(z_i\right)+\sigma_{i2}f_2\left(z_i\right)$$

当$f_1\left(z_i\right)f_2\left(z_i\right)\geq 0$时，

$$\sigma_i\left(f_1\left(z_i\right)+f_2\left(z_i\right)\right)=\sigma_{i1}f_1\left(z_i\right)+\sigma_{i2}f_2\left(z_i\right)$$

所以：

$$\hat{R}_Z\left(F_1+F_2\right)\leq\hat{R}_Z\left(F_1\right)+\hat{R}_Z\left(F_2\right)$$

也即：

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