CES 417T - HW4

467261 - Yifu Wang

2018 - 11 - 04

1

2 LFD Problem 4.25

(a) No, all the models are not validated on the same validation set. The VC-bound will change.

(b) In this case, the validation model is non-related to the validation set. So it's equivalent to select model from hypothesis applied in $\mathcal{D}_{val}$.

(c) Denote $m_i | i = 1, 2, 3...M$ for the ith modle. And K_i fit the Hoeffding inequality

$$\mathbb{P}[E_{out}(m_i) - E_{val}(m_i) > \epsilon] \leq e^{-\epsilon^2 K_i}$$

Then we have

$$\mathbb{P}[E_{out}(m^*) - E_{val}(m^*) > \epsilon] \leq \sum_{i=1}^{M} \mathbb{P}[E_{out}(m_i) - E_{val}(m_i) > \epsilon]$$

$$\leq \sum_{i=1}^{M} e^{-\epsilon^2 K_i}$$

Set

$$\kappa = -\frac{1}{2\epsilon^2} ln \left(\frac{1}{M} \sum_{i=1}^{M} e^{-\epsilon^2 K_i} \right)$$

We have

$$\mathbb{P}[E_{out}(m^*) - E_{val}(m^*) > \epsilon] \leq \sum_{i=1}^{M} e^{-\epsilon^2 K_i}$$

$$\leq M e^{ln\left(\frac{1}{M} \sum_{i=1}^{M} e^{-\epsilon^2 K_i} \right)}$$

$$\leq M e^{-\epsilon^2 \kappa(\epsilon)}$$

3 LFD Problem 5.4

(a)(i) We shouldn't use $N = 12500$ since that is the total data number of 50 years and we only use today's data. Or we shouldn't use $M = 500$ same for same reason.

(a)(ii) $\mathbb{P}[|E_{in} - E_{out}| > 0.02] \leq 2 \times 50000 \times e^{2 \times 12500 \times 0.02^2} \simeq 4.539993$

(b)(i) For the same reason stated in (a), we couldn't say buy and hold is a generally good choice since we get this concultion by only observing 500 stocks.

(b)(ii) If given the whole data sets and analyze the whole 50000 stocks maybe we can.