Introduction to Statistical Learning
Chapter 5 Exercise

Penghao Chen

March 16, 2019

1. This is an optimization problem. First we drive the target function:

$$Var(\alpha X + (1-\alpha)Y) = \alpha^2\sigma_X^2 + (1-\alpha)^2\sigma_Y^2 + 2\alpha(1-\alpha)\sigma_{XY}^2$$

Then we use first-order condition:

$$\begin{aligned}\frac{\partial Var}{\partial \alpha} &= 2\left(\alpha(\sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY}^2) - \sigma_Y^2 + \sigma_{XY}^2\right) \\ &= 0 \\ &\Rightarrow \alpha = \frac{\sigma_Y^2 - \sigma_{XY}^2}{\sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY}^2}\end{aligned}$$

We confirm this α leads to minimum rather than maximum point using second-order condition:

$$\begin{aligned}\frac{\partial^2 Var}{\partial \alpha^2} &= 2(\sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY}^2) \\ &= 2Var(X-Y) \\ &\geq 0\end{aligned}$$

The statement is proved.

2.

(a) $\frac{n-1}{n} = 1 - \frac{1}{n}$

(b) $\frac{n-1}{n} = 1 - \frac{1}{n}$

(c) With replacement, each draw from the original sample is independent. There are n draws, so the probability of not drawing the jth element for n times is $(1-\frac{1}{n})^n$

(d) $1 - (1-\frac{1}{5})^5 \approx 0.67$

(e) $1 - (1-\frac{1}{100})^100 \approx 0.63$

(f) $1 - (1-\frac{1}{10000})^10000 \approx 0.63$

(g) See code.

(h) See code.

3.

(a) Step 1, separate the original sample in to K non-overlapping sub-samples.

Step 2, select one sub-sample as the validation sample. Train the model on the other K-1 sub-samples. Apply the model onto the validation sample and collect prediction errors.

Step 3, repeat Step 2 and rotate the validation sample until all K sub-samples have been the validation sample once.

(b)

i The validation set approach has two major drawbacks:

1) The estimate of test error can be highly variable, depending on the choice of training and validation set.

2) This approach tends to overestimate the test error compared with the model fitted to the entire training set.

ii LOOCV also has two major drawbacks:

1) LOOCV is much more computationally intensive than K-fold cross-validation.

2) LOOCV test error has higher variance than K-fold validation test errors.

4. Using the bootstrap method:

Step 1, from the original sample, draw a sample of size n. Train the model, and get an estimate $\hat{Y}$ for X.

Step 2, repeat Step 1 for B times. Obtain the standard deviation of $\hat{Y}$ using formula (5.8).