5.1. $\alpha = \dfrac{\sigma_Y^2 - \sigma_{XY}}{\sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY}}$

α는 $Var(\alpha X + (1-\alpha)Y)$를

최소화

$$Var(\alpha X + (1-\alpha)Y) = Var(\alpha X) + Var((1-\alpha)Y) + 2Cov(\alpha X, (1-\alpha)Y)$$

$$= \alpha^2 Var(X) + (1-\alpha)^2 Var(Y) + 2\alpha(1-\alpha)Cov(X,Y)$$

양변을 α에 대해
미분하면

$$0 = 2\alpha Var(X) - 2(1-\alpha)Var(Y) + (2-4\alpha)Cov(X,Y)$$

α에 대해 정리하면 5.6과 동일한 식이 된다.

5.2. n개 관측치

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

모든 관측치: 뽑힐 확률이 같음.
j번째 제외한 개수 → $n-1$

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

(a)와 비슷한 논리: n개 중에서 원래 표본의
j번째를 제외한 것을 뽑을 확률 → $\frac{(n-1)}{n}$

(c) j번째 관측치가 붓스트랩 표본에 있지 않을 확률

$$\underbrace{\left(\frac{n-1}{n}\right) \times \left(\frac{n-1}{n}\right) \times \cdots \left(\frac{n-1}{n}\right)}_{n\text{개에 대해서 위 작업 수행}} = \left(\frac{n-1}{n}\right)^n$$

(d) $n=5$일때 j번째 관측치가 붓스트랩 표본에
있을 확률

$$1-\left(\frac{5-1}{5}\right)^5$$

(e) 〃 100 $\rightarrow 1-\left(1-\frac{1}{100}\right)^{100}$

(f) 〃 10000

$$1-\left(1-\frac{1}{10000}\right)^{10000}$$

(g), (h)는 코드에서 〃

5.3

(a) 전체 데이터셋을 k개로 나누고
k-1개는 training set,
1개는 test set로 분류하고
교차검증 k번 반복 후 평균도출

(b) 검증셋 : +) 구현이 쉽다
-) 추정치가 어느 셋이 훈련/검증셋에 포함되냐에 따라 변동이 생긴다.

LOOCV : +) 검증셋의 단점을 갖지 않는다.
-) 계산식이 지나치게 복잡할 수 있다!

5.4. 예측값의 표준편차를 어떻게 추정하는가?

부스트랩 과정을 여러 번 반복하고 이로 인해 구한 데이터셋을 통해 모델을 만들어 예측한다.