Exercise Sheet 7

2. Dezember 2018

1 Bias and Variance of Mean Estimators

Calculation of bias, variance and mean squared error of the following estimators:

(a) $\hat{\mu} = \frac{1}{N}\sum_i^N X_i$

$$Bias(\hat{\mu}) = E[\hat{\mu} - \mu] = E[\frac{1}{N}\sum_i^N X_i - \mu] = \frac{1}{N}E[\sum_i^N X_i] - \mu = \frac{1}{N} \cdot N \cdot \mu - \mu = 0$$

$$Var(\hat{\mu}) = E[(\hat{\mu} - E[\hat{\mu}])^2] = \frac{1}{N^2}\sum_i^N Var(X_i) = \frac{1}{N}\mu$$

$$MSE(\hat{\mu}) = Var(\hat{\mu}) + Bias^2(\hat{\mu}) = Var(\hat{\mu}) = \frac{1}{N}\mu$$

(b) $\hat{\mu} = 0$

$$Bias(\hat{\mu}) = E[\hat{\mu} - \mu] = E[0 - \mu] = -\mu$$
$$Var(\hat{\mu}) = E[(\hat{\mu} - E[\hat{\mu}])^2] = E[(0 - E[0])^2] = 0$$
$$MSE(\hat{\mu}) = Var(\hat{\mu}) + Bias^2(\hat{\mu}) = Bias^2(\hat{\mu}) = \mu^2$$

2 Bias-Variance Decomposition for Regression

(a) Proof of $Error(\hat{f}(x)) = Bias(\hat{f}(x))^2 + Var(\hat{f}(x))$

$$Error(\hat{f}(x)) = E[(\hat{f}(x) - f(x))^2]$$

Adding Zero doesn't change the equation

$$= E[((\hat{f}(x) - E[\hat{f}(x)]) + (E[\hat{f}(x)] - f(x)))^2]$$

Multiplicating out the square

$$= E[(\hat{f}(x) - E[\hat{f}(x)])^2 + 2(\hat{f}(x) - E[\hat{f}(x)])(E[\hat{f}(x)] - f(x)) + (E[\hat{f}(x)] - f(x))^2]$$

And because $E[\hat{f}(x) - E[\hat{f}(x)]] = 0$ and in the squareterms we are allowed to change the summands, we obtain,

$$= E[(\hat{f}(x) - E[\hat{f}(x)])^2] + E[(E[\hat{f}(x)] - f(x))^2] = Var(\hat{f}(x)) + Bias(\hat{f}(x))^2$$

3 Bias-Variance Decomposition for Classification

(a) Solution for the Optimisation problem $min_R = E[D_{KL}(P|\hat{P})]$

$$E[D_{KL}(P|\hat{P})] = E[\sum_i P_i log(\frac{P_i}{\hat{P}_i})]$$

for a minimization problem we use the Lagrange-method

$$L(P,\lambda) = \sum_i P_i log(\frac{P_i}{\hat{P}_i}) + \lambda(\sum_i P_i - 1)$$

obtaining as partial derivatives

$$\frac{\partial L}{\partial P} = \sum_i (\frac{P_i}{\hat{P}_i} + 1) + \lambda = 0, \quad \frac{\partial L}{\partial \lambda} = \sum_i P_i - 1 = 0$$

Following the term for P_i

$$P_i = \frac{exp(E[log(\hat{P}_i)])}{\sum_j exp(E[log(\hat{P}_j)])}$$

And with $R_i = P_i$ we obtain $R = [R_1, ..., R_C]$ as solution.

(b) Proof of $Error(\hat{P}) = Bias(\hat{P}) + Var(\hat{P})$

$$Bias(\hat{P}) = \sum_i P_i log(\frac{P_i}{R_i}) = \sum_i P_i log(\frac{P_i}{\frac{exp(E[log(\hat{P}_i)])}{\sum_j exp(E[log(\hat{P}_j)])}} = \sum_i P_i log(\frac{P_i}{\hat{P}_i}) + log(\sum_j exp(E[log(\hat{P}_j)]))$$

In (a) we showed the R that is minimizing the expected divergence and under usage of $\sum_i R_i = 1$

$$log(\sum_j exp(E[log(\hat{P}_j)])) = log(\frac{exp(E[log(\hat{P}_i)])}{R_i} = E[\sum_i R_i log(\frac{\hat{P}_i}{R_i})] = -E[D_{KL}(R|\hat{P})] = -Var(\hat{P})$$

Putting our setup into the Term for the Bias

$$Bias(\hat{P}) = D_{KL}(P|R) = E[D_{KL}(P|\hat{P})] - E[D_{KL}(R|\hat{P})]$$

So that in the end we obtain

$$Error(\hat{P}) = E[D_{KL}(P|\hat{P})] = D_{KL}(P|R) + E[D_{KL}(R|\hat{P})] = Bias(\hat{P}) + Var(\hat{P})$$