

EXERCISE SET 7

Exercise 1 (Mean-squared error): Prove that

$$\widetilde{\text{MSE}}_{\hat{y}(x,D)} = \sigma^2 + \text{var}(\hat{y}(x, D)) + E [\hat{y}(x, D) - f(x)]^2$$

where $\widetilde{\text{MSE}}_{\hat{y}(x,D)}$ is defined in Equation 3 in the class notes.

Exercise 2 (Laplace's rule): Consider a Bernoulli random variable

$$X \sim \text{Bernoulli}(q).$$

(You can assume this means $P(X = 1) = q$). If we have samples X_1, \dots, X_N for X then we have seen that a consistent and unbiased estimator of q is

$$\hat{q} = \frac{Y}{N}, \text{ where } Y = \sum_{i=1}^N X_i.$$

An alternative estimator, called Laplace's rule of succession, is

$$\hat{q}_L = \frac{Y + 1}{N + 2}.$$

The motivation of for defining \hat{q}_L is as follows: Think of X as a biased coin. If we know that it is possible for to roll a heads or a tails, then we should include this information in our estimator. However, using the original estimator, if we roll a sequence of only heads (or tails), we will estimate that $q = 1$ (or $q = 0$). To correct for this, we pretend we have two additional observations (hence the $N + 2$ in the denominator) and that one is heads and one is tails (hence $Y + 1$ in the numerator). This is a simple example where we are incorporating *prior* information into our estimator – that is, information beyond what is present in the data and our model. In this case, that prior information is that the coin has two sides and could land on either one, however unlikely that might be.

- (a) Derive formula for the mean-squared error,

$$\text{MSE}_{\hat{q}_L} = E [(\hat{q} - q)^2]$$

and decompose it into the variance and the squared bias. Your formulas should be in terms of q and N .

- (b) Is \hat{q}_L unbiased and consistent?

- (c) Now compute $\text{MSE}_{\hat{q}}$. Note that in this case the bias is zero, so it should be straightforward to derive from the standard error. Surprisingly, $\text{MSE}_{\hat{q}} > \text{MSE}_{\hat{q}_L}$ for some values of N and q . For which values is this the case? Note that this is quite surprising, since it seems like \hat{q} should be the best guess of q !

Exercise 3 (Impulse function features): Consider the model

$$Y|X \sim \text{Normal} \left(\sum_{i=1}^K \beta_i \phi_i(X), \sigma^2 \right)$$

Suppose that $X \in [0, 1)$ and define the intervals

$$I_i = \left[\frac{i-1}{K}, \frac{i}{K} \right)$$

Notice that

$$[0, 1) = I_1 \cup I_2 \cup \dots \cup I_K.$$

That is, each x in $[0, 1)$ is in one of these disjoint intervals. Now introduce the features

$$\phi_i(x) = \begin{cases} 1 & x \in [(i-1)/K, i/K) \\ 0 & x \notin [(i-1)/K, i/K) \end{cases}$$

- (a) Are ϕ_i orthogonal with respect to a random variable X taking values in $[0, 1)$? Does it depend on the distribution of the random variable?
- (b) Using `statsmodels`, implement fitting the model with these features. You can make up your simulated data set to fit, or copy the code I used in class to fit the fourier and polynomial models. I recommend writing a function `phi(x,i)` which takes the array of predictors and outputs an array $[\phi_1(X_1), \dots, \phi_j(X_N)]$. Use $K = 10$ and $N = 100$.
- (c) As usual let $\hat{\beta}_j$ be the fitted value of β_j using least squares, meaning the value that minimizes the squared residuals. Show that in this model $\hat{\beta}_j$ is simply the average value of Y_i among data points where $X_i \in I_j$; that is,

$$\hat{\beta}_j = \frac{1}{N_j} \sum_{i: X_i \in I_j} Y_i$$

where N_j is the number of points in I_j .