

Exercise 2.7 ESL

January 30th 2026

Setup

1 Setup

2 Exercise (a)

- Linear regression
- k -NN

3 Exercise (b)

4 Exercise (c)

5 Exercise (d)

Setup

We have a sample of N pairs (x_i, y_i) drawn i.i.d. from the following distribution:

$$\begin{array}{ll} x_i \sim h(x), & \text{the design density} \\ y_i = f(x_i) + \varepsilon_i, & f \text{ is the regression function} \\ \varepsilon_i \sim (0, \sigma^2) & (\text{mean } 0, \text{variance } \sigma^2) \end{array}$$

Consider estimators for f linear in the y_i :

$$\hat{f}(x_0) = \sum_{i=1}^n \ell_i(x_0; \mathcal{X}) y_i, \quad (1)$$

where the weights $\ell_i(x_0; \mathcal{X})$ do not depend on the y_i , but depend on the entire training sequence of x_i , here denoted by $\mathcal{X}$.

Exercise (a)

1 Setup

2 Exercise (a)

- Linear regression
- k -NN

3 Exercise (b)

4 Exercise (c)

5 Exercise (d)

Exercise (a): linear regression

Show that the linear regression estimator belongs to the class of estimators that can be written on the form

$$\hat{f}(\mathbf{x}_0) = \sum_{i=1}^n \ell_i(\mathbf{x}_0; \mathcal{X}) y_i.$$

Solution

Response vector: $\mathbf{y} = (y_1, \dots, y_N)^T \in \mathbb{R}^{N \times 1}$

Design matrix: $\mathbf{X} \in \mathbb{R}^{N \times p}$

When $\mathbf{X}$ has full rank, the least squares estimator is given by¹:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}.$$

The estimator for f can be written as

$$\hat{f}(\mathbf{x}_0) = \mathbf{x}_0^T \hat{\boldsymbol{\beta}}.$$

Note: For two column vectors $\mathbf{a}, \mathbf{b}$, each of length p , the operation $\mathbf{a}^T \mathbf{b}$ is the same as the scalar product. That is, $\mathbf{a}^T \mathbf{b} = \mathbf{a} \cdot \mathbf{b} = a_1 b_1 + a_2 b_2 + \dots + a_p b_p$.

¹See page 12 in the textbook.

We get that

$$\begin{aligned}\hat{f}(\mathbf{x}_0) &= \mathbf{x}_0^T \hat{\boldsymbol{\beta}} \\ &= \mathbf{x}_0^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \\ &= (\mathbf{x}_0^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T) \mathbf{y} \\ &= (\mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_0)^T \mathbf{y} \\ &= \sum_{i=1}^N (\mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_0)_i y_i \\ &= \sum_{i=1}^N \ell_i(\mathbf{x}_0; \mathcal{X}) y_i,\end{aligned}$$

where we have let $\ell_i(\mathbf{x}_0; \mathcal{X}) = (\mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_0)_i$.

Exercise (a): k -NN

We redo the previous exercise for the k -NN estimator²:

$$\hat{f}(x_0) = \frac{1}{k} \sum_{x_i \in N_k(x_0)} y_i$$

²See page 14 in the textbook.

We get that

$$\begin{aligned}\hat{f}(x_0) &= \frac{1}{k} \sum_{x_i \in N_k(x_0)} y_i \\ &= \frac{1}{k} \sum_{i=1}^N 1(x_i \in N_k(x_0)) y_i \\ &= \sum_{i=1}^N \frac{1}{k} 1(x_i \in N_k(x_0)) y_i \\ &= \sum_{i=1}^N \ell_i(x_0; \mathcal{X}) y_i,\end{aligned}$$

where we have let $\ell_i(x_0; \mathcal{X}) = \frac{1}{k} 1(x_i \in N_k(x_0))$.

Exercise (b)

1 Setup

2 Exercise (a)

- Linear regression
- k -NN

3 Exercise (b)

4 Exercise (c)

5 Exercise (d)

Exercise (b)

Derive the conditional bias-variance decomposition:

$$E_{\mathcal{Y}|\mathcal{X}}[(f(x_0) - \hat{f}(x_0))^2] = \text{Bias}_{\mathcal{Y}|\mathcal{X}}(\hat{f}(x_0))^2 + \text{Var}_{\mathcal{Y}|\mathcal{X}}(\hat{f}(x_0)).$$

Note that

Inside the expectation, the training sequence $\mathcal{X}$ is fixed.

We treat x_0 and $f(x_0)$ as fixed quantities.

Since we condition on $\mathcal{X}$, the only random part is $\hat{f}(x_0)$, as it depends on $\mathcal{Y}$ (see the Setup slide).

Solution

Use the add-and-subtract trick:

$$\begin{aligned} & \mathbb{E}_{\mathcal{Y}|\mathcal{X}} \left[\left(f(x_0) - \hat{f}(x_0) \right)^2 \right] \\ &= \mathbb{E}_{\mathcal{Y}|\mathcal{X}} \left[\left(f(x_0) - \mathbb{E}_{\mathcal{Y}|\mathcal{X}}[\hat{f}(x_0)] + \mathbb{E}_{\mathcal{Y}|\mathcal{X}}[\hat{f}(x_0)] - \hat{f}(x_0) \right)^2 \right] \\ &= \mathbb{E}_{\mathcal{Y}|\mathcal{X}} \left[\left(\left(f(x_0) - \mathbb{E}_{\mathcal{Y}|\mathcal{X}}[\hat{f}(x_0)] \right) + \left(\mathbb{E}_{\mathcal{Y}|\mathcal{X}}[\hat{f}(x_0)] - \hat{f}(x_0) \right) \right)^2 \right] \\ &= \mathbb{E}_{\mathcal{Y}|\mathcal{X}} \left[\left(f(x_0) - \mathbb{E}_{\mathcal{Y}|\mathcal{X}}[\hat{f}(x_0)] \right)^2 \right] \\ &+ \mathbb{E}_{\mathcal{Y}|\mathcal{X}} \left[2 \cdot \left(f(x_0) - \mathbb{E}_{\mathcal{Y}|\mathcal{X}}[\hat{f}(x_0)] \right) \left(\mathbb{E}_{\mathcal{Y}|\mathcal{X}}[\hat{f}(x_0)] - \hat{f}(x_0) \right) \right] \\ &+ \mathbb{E}_{\mathcal{Y}|\mathcal{X}} \left[\left(\mathbb{E}_{\mathcal{Y}|\mathcal{X}}[\hat{f}(x_0)] - \hat{f}(x_0) \right)^2 \right]. \end{aligned}$$

$$\begin{aligned} \mathbb{E}_{\mathcal{Y}|\mathcal{X}} \left[\left(f(x_0) - \mathbb{E}_{\mathcal{Y}|\mathcal{X}}(\hat{f}(x_0)) \right)^2 \right] &= \left(f(x_0) - \mathbb{E}_{\mathcal{Y}|\mathcal{X}}(\hat{f}(x_0)) \right)^2 \\ &= \text{Bias}_{\mathcal{Y}|\mathcal{X}}(\hat{f}(x_0))^2. \end{aligned}$$

$$\begin{aligned} & 2 \cdot \mathbb{E}_{\mathcal{Y}|\mathcal{X}} \left[(f(x_0) - \mathbb{E}_{\mathcal{Y}|\mathcal{X}} \hat{f}(x_0)) (\mathbb{E}_{\mathcal{Y}|\mathcal{X}} \hat{f}(x_0) - \hat{f}(x_0)) \right] \\ &= 2 \cdot (f(x_0) - \mathbb{E}_{\mathcal{Y}|\mathcal{X}} \hat{f}(x_0)) \cdot \mathbb{E}_{\mathcal{Y}|\mathcal{X}} \left[\mathbb{E}_{\mathcal{Y}|\mathcal{X}} \hat{f}(x_0) - \hat{f}(x_0) \right] \\ &= 2 \cdot (f(x_0) - \mathbb{E}_{\mathcal{Y}|\mathcal{X}} \hat{f}(x_0)) \cdot (\mathbb{E}_{\mathcal{Y}|\mathcal{X}} \hat{f}(x_0) - \mathbb{E}_{\mathcal{Y}|\mathcal{X}} \hat{f}(x_0)) \\ &= 0. \end{aligned}$$

$$E_{\mathcal{Y}|\mathcal{X}}[(E_{\mathcal{Y}|\mathcal{X}}\hat{f}(x_0) - \hat{f}(x_0))^2] = \text{Var}_{\mathcal{Y}|\mathcal{X}}(\hat{f}(x_0)).$$

Combining our results

We get that

$$E_{\mathcal{Y}|\mathcal{X}}[(f(x_0) - \hat{f}(x_0))^2] = \text{Bias}_{\mathcal{Y}|\mathcal{X}}(\hat{f}(x_0))^2 + \text{Var}_{\mathcal{Y}|\mathcal{X}}(\hat{f}(x_0)),$$

which was what we wanted to show.

Exercise (c)

1 Setup

2 Exercise (a)

- Linear regression
- k -NN

3 Exercise (b)

4 Exercise (c)

5 Exercise (d)

Exercise (c)

Derive the unconditional bias-variance composition:

$$\mathbb{E}_{\mathcal{X}, \mathcal{Y}} \left[(f(x_0) - \hat{f}(x_0))^2 \right].$$

In this case

We treat both training sequences $\mathcal{X}$ and $\mathcal{Y}$ as random.

We still treat x_0 and $f(x_0)$ as fixed quantities

Repeat the steps as in the previous exercise but with the unconditional expectation,

$$E_{\mathcal{X}, \mathcal{Y}}[(f(x_0) - \hat{f}(x_0))^2].$$

Exercise (d)

1 Setup

2 Exercise (a)

- Linear regression
- k -NN

3 Exercise (b)

4 Exercise (c)

5 Exercise (d)

Exercise (d)

Establish a relationship between the squared biases and variances in the above two cases.

Recall the law of total expectation:

$$\mathbb{E}_{\mathcal{X}, \mathcal{Y}}[(f(x_0) - \hat{f}(x_0))^2] = \mathbb{E}_{\mathcal{X}}[\mathbb{E}_{\mathcal{Y}|\mathcal{X}}[(f(x_0) - \hat{f}(x_0))^2]].$$

From this, we obtain the relationship

$$\text{Var}(\hat{f}(x_0)) + (\text{Bias}(\hat{f}(x_0)))^2 = \mathbb{E}_{\mathcal{X}}[\text{Var}_{\mathcal{Y}|\mathcal{X}}(\hat{f}(x_0))] + \mathbb{E}_{\mathcal{X}}[(\text{Bias}_{\mathcal{Y}|\mathcal{X}}(\hat{f}(x_0)))^2].$$