

Compulsory exercise 1: Group 5

TMA4268 Statistical Learning V2022

Aleksander Johnsen Solberg and Gjermund Oscar Lyckander

08 February, 2022

Problem 1

a)

$$\begin{aligned} E[y_0 - \hat{f}(x_0)]^2 &= E[(f(x_0) + \varepsilon - \hat{f}(x_0))^2] \\ &= E[(f(x_0))^2] + E[\varepsilon^2] + E[\hat{f}(x_0)^2] - 2E[f(x_0)\hat{f}(x_0)] + 2E[f(x_0)\varepsilon] + 2E[\hat{f}(x_0)\varepsilon] \\ &= f(x_0)^2 + \text{Var}(\varepsilon) + \text{Var}(\hat{f}(x_0)) + E[\hat{f}(x_0)]^2 - 2E[f(x_0)\hat{f}(x_0)] \\ &= (f(x_0) - E[\hat{f}(x_0)])^2 + \text{Var}(\hat{f}(x_0)) + \text{Var}(\varepsilon) \end{aligned}$$

The three terms in the last line are the squared bias, variance, and irreducible error respectively.

b)

The three terms can be interpreted as the following. The bias term is the error that comes from modeling a complicated real-life problem with a simple model. The more flexible the model is, the smaller the bias will be. The variance term is how much the estimate \hat{f} would change if we were using different training data. The more flexible the model is, the larger the variance will be. Lastly, the irreducible error term is simply the error that comes from the error in the data itself.

c)

d)

e)

Problem 2

Here is a code chunk:

```
library(palmerpenguins) # Contains the data set "penguins".
data(penguins)
head(penguins)

## # A tibble: 6 x 8
##   species island bill_length_mm bill_depth_mm flipper_length_~ body_mass_g sex
##   <fct>   <fct>         <dbl>         <dbl>         <int>     <int> <fct>
## 1 Adelie  Torge~         39.1         18.7         181       3750 male
## 2 Adelie  Torge~         39.5         17.4         186       3800 fema~
## 3 Adelie  Torge~         40.3          18         195       3250 fema~
## 4 Adelie  Torge~          NA          NA          NA         NA NA
## 5 Adelie  Torge~         36.7         19.3         193       3450 fema~
```

```
## 6 Adelie Torge~      39.3      20.6      190      3650 male
## # ... with 1 more variable: year <int>
```

a)

b)

c)

Problem 3

Problem 4