Compulsory exercise 1: Group 16

TMA4268 Statistical Learning V2022

Weicheng Hua, Emil Johannese Haugstvedt, Torbjørn Baadsvik

17 February, 2022

Problem 1

a)

The expected MSE on the test set is given by:

$$\begin{aligned}
E[(y_0 - \hat{f}(x_0))^2] &= E[(f(x_0) - \hat{f}(x_0) + \epsilon)^2] \\
&= E[(f(x_0) - \hat{f}(x_0))^2] + 2E[\epsilon(f(x_0) - \hat{f}(x_0))] + E[\epsilon^2] \\
&= E[f(x_0)^2 - 2f(x_0)\hat{f}(x_0)] + E[\hat{f}(x_0)^2] + E[\epsilon^2] \\
&= E[f(x_0)^2 - 2f(x_0)\hat{f}(x_0) + \hat{f}(x_0)^2] + (E[\hat{f}(x_0)^2] - E[\hat{f}(x_0)]^2) + E[\epsilon^2] \\
&= E[(f(x_0) - \hat{f}(x_0))^2] + (E[\hat{f}(x_0)^2] - E[\hat{f}(x_0)]^2) + E[\epsilon^2] \\
&= E[(f(x_0) - \hat{f}(x_0))^2] + Var[\hat{f}(x_0)] + Var[\epsilon] \\
&= \text{Squared bias} + \text{Variance of prediction} + \text{Irreducible error}
\end{aligned}$$

b)

The squared bias term represents the expected squared deviation between the prediction of the "true" model and the prediction of the fitted model. The variance of prediction term represents the degree to which the prediction of the fitted model can vary depending on the input. Higher variance of prediction means the model can adapt it's prediction to input data to a greater extent than a simpler model, implying that the model is more flexible. However, the increased "adaptability" may be unwanted if it leads to overfitting.

c)

i	ii	iii	iv
$TRUE$	$FALSE$	$TRUE$	$FALSE$

d)

i	ii	iii	iv
$TRUE$	$FALSE$	$TRUE$	$FALSE$

e)

```
library(matrixcalc)
mat <- cbind(c(50, 33, 18), c(33, 38, -10), c(18, -10, 72))
is.positive.semi.definite(mat)
```

```
## [1] TRUE
```

Answer: iii) 0.76

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