Compulsory exercise 1: Group 21

TMA4268 Statistical Learning V2018

Jørgen Opheim, Ole-Andreas Sandnes and Sander Coates

[Date]

Take a look at the cheat sheets for R Markdown here: File > Help > Cheatsheets > R Markdown Cheat Sheet in RStudio, or here http://www.rstudio.com/wp-content/uploads/2016/03/rmarkdown-cheatsheet-2.0.pdf

or the lessons: http://rmarkdown.rstudio.com/lesson-1.html

Problem 1 - Core concepts in statistical learning [2 points]

a) Training and test MSE [1 point]

Here you write your smart answer. You may write latex $Y = f(x) + \varepsilon$.

Items with

- item1
- item2

b) Bias-variance trade-off [1 point]

- Explain how that is done. Hint: this is what the M repeated training data sets are used for.
- Focus on Figure 4. As the flexibility of the model increases (K decreases), what happens with
  - the squared bias,
  - the variance, and
  - the irreducible error?
- What would you recommend is the optimal value of K? Is this in agreement with what you found in a)?

Problem 2 - Linear regression [4 points]

Here you see an R chunk that is evaluated (when knitting) and code is displayed.

```
library(ggplot2)
data = read.table("https://www.math.ntnu.no/emner/TMA4268/2018v/data/SYSBPreg3uid.txt")
dim(data)
colnames(data)
modelA=lm(-1/sqrt(SYSBP) ~ .,data = data)
summary(modelA)
```

a) Understanding model output [1 point]

b) Model fit [1 point]

c) Confidence interval and hypothesis test [1 points]

d) Prediction [1 point]