The exam

Gerko Vink

Fundamental Techniques in Data Science with R

What can be tested

The information in the lecture slides:

- An archive of all lecture slides. Or as seperate lectures:
 - Lecture 1
 - Lecture 2
 - Lecture 3
 - Lecture 4
 - Lecture 5
 - Lecture 6
 - Lecture 7
 - Lecture 8

and the information in the following sources these lecture slides are based on:

- Introduction to Statistical Learning
 - Chapter 2: up to and including 2.2.2
 - Chapter 3: up to and including 3.4
 - Chapter 4: up to and including 4.3.4
 - Chapter 5: up to and including 5.1
- These lecture notes.
- This document I prepared about interpreting cross validation results
- This online page that details the relation between correlation and prediction
- This blog post by Jonathan Barlett that details R^2 and explains how it is useless on a wrongly specified • model
- This other blog post by Jonathan about adjusted R^2
- This page on interpreting logistic regression
- This page on inspecting logistic regression

What about equations and formulae?

Your knowledge of matrix algebra will not be tested. So, there is no need to memorize that the regression estimates β can be estimated as $\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T y$. However, you will need to know, understand and apply equations such as:

- $y = \beta_0 + \beta_1 X + \epsilon$ and any more complicated version of this.
- $\epsilon = y \hat{y}$
- $\mathbb{E}[y] = \alpha + \beta x.$
- $\log(\operatorname{odds}) = \log(\frac{p}{1-p}) = \log(p) \log(1-p) = \log(p)$ $p_i = \frac{\exp(\eta)}{1+\exp(\eta)} = \frac{\exp(\beta_0 + \beta_1 x_{1,i} + \dots + \beta_n x_{n,i})}{1+\exp(\beta_0 + \beta_1 x_{1,i} + \dots + \beta_n x_{n,i})}$

- etcetera

What if you are in doubt?

If any of the course materials confuse you, drop me a line and I'd be more than happy to explain. The second half of the last lecture is dedicated to a Q&A

End of document