# Decision Support Systems
# Team 3

Repport
Aarhus University, Science and Technology
Lector: Christian Fischer Pedersen

March 1, 2018

| **Name** | **Study number** | **Signature** |
|---|---|---|
| David Jensen | 11229 | |
| Henrik Bagger Jensen | 201304157 | |
| Ólafur Dagur Skúlason | IY11249 | |
| Titas Urbonas | 201700321 | |
| Christian M. Lillelund | 201408354 | |

# Contents

# Chapter 1: Introduction

# Chapter 2: Regression

This chapter details the work of LAB exercise 3.6.2, 3.6.3, 4.6.1 and 4.6.2 from "An Introduction to Statistical Learning". It starts by recapitulating the theory behind linear regressions, both simple and multiple, then proceeds to describe the accompanied LAB exercises and conclude on their findings.

## 2.1 Multiple Linear Regression

### 2.1.1 Theory

Basic theory for simple and multiple lin regs here. From the slides or book[1].

Simple Linear Regression is used to make linear models of data. It has a response Y on the basis of a single predictor variable X. We can write it as:

$$Y = \beta_0 + \beta_1 X_1 + \epsilon_i \tag{2.1}$$

$\beta_0 + \beta_1$ are unknown and to get a response, we must use data to estimate the coefficients.$(x_1, y_1)$, $(x_2, y_2)$,..., $(x_n, y_n)$ represent n observation pairs, each of which consists of a measurement of X and a measurement of Y. The drawback of this method is that only a single predictor variable is used and often have more. In cases where we want examined the relationship between multiple predictor variables we use Multiple Linear Regression. The model takes the following form:

$$Y = \beta_0 + \beta_1 X_1 + \ldots + \beta_n X_n + \epsilon_i \tag{2.2}$$

To obtain the estimated Coefficients in the model we use the least squares method to minimize the sum of squared residuals. We pick $\beta_0, \beta_1, ...\beta_p$ to to minimize the sum of squared residuals.

$$RSS = \sum (y - \hat{y})^2 = \sum (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \ldots - \hat{\beta}_p x_i p)^2 \tag{2.3}$$

To evaluated the model we can use RSE (residual standard error). This is done by meaning and rooting the result of the RSS. The outcome of the formula is the average amount that the response will deviate from the regression line. This is also known as an estimate of the $\epsilon$ in the Standard Linear Regression formula (2.1) stated earlier in this chapter:

$$RSE = \sqrt{\frac{1}{n-2} \cdot RSS} \tag{2.4}$$

[1] [James et al.(2013)James, Witten, Hastie, and Tibshirani]

### 2.1.2 Results

#### LAB 3.6.2

#### LAB 3.6.3

### 2.1.3 Conclusion

### 2.1.4 Logistic Regression

### 2.1.5 Theory

Basic theory for logistic lin regs here. From the slides or book.

### 2.1.6 Results

LAB 4.6.1 + 4.6.2

### 2.1.7 Conclusion

# Chapter 3: Linear Discriminant Analysis

# Chapter 4: Cross Validation

# Chapter 5: Subset Selection

# Chapter 6: Shrinkage Methods

# Chapter 7: Clustering Methods

# Chapter 8: Discussion

# Chapter 9: Conclusion

# Chapter 10: Perspectives

# Bibliography

[James et al.(2013)James, Witten, Hastie, and Tibshirani] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An Introduction to Statistical Learning*, volume 103 of *Springer Texts in Statistics*. Springer New York, New York, NY, 2013. ISBN 9781461471387. doi: 10.1007/978-1-4614-7138-7. URL `http://www-bcf.usc.edu/{~}gareth/ISL/`.