Week 1

# Assignment 0

Norman Rzepka, 754644

Markus Dücker, 779867

# Assignment 1

## a)

mean: 0.508

median: 0.545

variance: 0.141596

## b)

mean: 1.068921

median: 1.029366

variance: 0.381223

## c)



# Assignment 2

## Eruptions:

min: 1.6

max: 5.1

mean: 3.487783

median: 4

abs diff of median and mean: 0.512217

## Waiting:

min: 43

max: 96

mean: 70.897059

median: 76

abs diff of median and mean: 5.102941

# Assignment 3

## a)

number of different species: 3

50 data entries each

## b)



It seems as if there were 2 clusters in the data (as shown below), which might help us to separate the classes.



## c)



It turns out one of the clusters is only related to Iris setosa. So our hypothesis was valid and we may use these 2 parameters to separate Iris setosa from the other classes.

## d)



(red: Iris setosa, blue: other Iris species)

# Assignment 4





# Assignment 5

Height = 0.614 \* Age + 66.633



# Assignment 6

1. High precision: **Medical diagnosis** (e.g. cancer detection based on medical metrics). Starting distressing and costly treatment for false positive patients may have fatal consequences.
2. High recall: **Quality Assurance**. Shipping a faulty product to a customer because of misclassification in the QA process may lead to brand damage of the producer or worse consequences. In some areas (e.g. car industry) a faulty product may even lead to fatal injuries.
3. High Accuracy: **Optical Character Recognition (OCR)**. There are many classes when doing OCR and it is favorable to detect all possible characters with a fairly high accuracy. There is no point in favoring one character over the other.

# Assignment 7

Example thresholds (predicted class is yes if predicted value is greater than or equal to the threshold):

1. High precision: 0.95
2. High recall: 0.48
3. High accuracy: 0.73, 0.63, 0.56, 0.48 or 0

# Bonus Assignment 1

