

Project 1  
Course 02445  
Project in Statistical evaluation of  
artificial intelligence

Rasmus J. P. s164564  
Nikolaj S. P. s183930

January 2020

# 1 Introduction

## 2 Data

Something general about the data

Person 9 is missing some of the initial 1-4 datapoints for some of the experiments (not experiment 2), we impute the values with the first available datapoints, such that the first 1-5 datapoints are the same for those particular repetitions. This seems reasonable since it is only a few datapoints and they are located at the beginning, so it is equivalent to the person starting from that position and holding it there for the first few measurements.

## 3 Method

### 3.1 Artificial Neural Network

A classification network was trained to classify a person from the curve created by their movement in experiment 2. Several architectures were trained to find the best suited architecture for the task, in the end we decided on this particular architecture:

Layer no.	Function
Layer 1	linear(300, 150) ReLU Dropout(0.15)
Layer 2	linear(150, 75) ReLU Dropout(0.15)
Layer 3	linear(75, 10)
Layer 4	Softmax

A leave one out cross-validation was then performed on the network, to estimate its generalization error.

### 3.2 Clustering

$$E = \sum_{i=1}^N \sum_{k=1}^K z_{ik} \|\mathbf{x}_i - \boldsymbol{\mu}_k\|_2^2 \quad (1)$$

$$z_{ik} = \frac{1}{K} \quad (2)$$

Notice how  $z_{ik}$  and  $\mu_k$  will change every time a cluster is updated. The formulae for K-means computes the sum of squares between each observation in a cluster to the cluster mean. Ward's method or Ward linkage compares the K-means error from any giving merger between two observation or clusters. Combination of two observations or clusters are then determined by the merger with lowest increase to the K-means

error. Ward linkage then becomes the function that determines the agglomeration that creates the following dendrogram.

$$\mu_1 = \begin{bmatrix} -0.67 \\ 0.78 \\ 1.12 \\ -0.45 \\ -0.85 \\ 1.18 \\ 1.27 \\ 0.6 \\ -0.8 \\ 1.11 \\ 1.23 \end{bmatrix} \quad \mu_2 = \begin{bmatrix} 0.33 \\ -0.39 \\ -0.56 \\ 0.22 \\ 0.42 \\ -0.59 \\ -0.63 \\ -0.3 \\ 0.4 \\ -0.55 \\ -0.61 \end{bmatrix} \quad \mu_6 = \begin{bmatrix} -1.36 \\ 2.28 \\ 1.13 \\ -0.58 \\ -1.04 \\ 1.66 \\ 1.53 \\ 0.8 \\ -1.89 \\ 1.15 \\ 1.44 \end{bmatrix} \quad (3)$$

$$\hat{\mu}_1 = \begin{bmatrix} 1.51 \\ 0.08 \end{bmatrix} \quad \hat{\mu}_2 = \begin{bmatrix} -3.16 \\ -0.17 \end{bmatrix} \quad \hat{\mu}_6 = \begin{bmatrix} -2.02 \\ 3.45 \end{bmatrix} \quad (4)$$

## Evaluating clusters

We use Rand index to evaluate the clustering of our different methods. Specifically we will compare our true labels with the labels given by a model. We define two measures from our clustering:

$$S = \sum_{i=1}^{N-1} \sum_{j=i+1}^N S_{ij} \quad (5)$$

$$D = \sum_{i=1}^{N-1} \sum_{j=i+1}^N D_{ij} \quad (6)$$

The notation requires some explanation. We define two cluster Z (true) and Q (predicted).  $S_{ij} = 1$  if and only if Z and Q agrees that the pair of observation  $x_i, x_j$  belong to the same cluster, otherwise  $S_{ij} = 0$ . Similar  $D_{ij} = 1$  only if Z and Q agrees that the pair of observation  $x_i, x_j$  doesn't belong to the same cluster. We then calculate the rand index like:

$$R(Q, P) = \frac{S + D}{\frac{1}{2}N(N-1)} \quad (7)$$