Functions

**Pre-processing**

Zscore\_omit\_nan: calculates the zscore of a matrix that contains NaN values, omitting the NaN

input

InterpolateNaN: interpolates outliers to NaN

* Input
  + Class Matrix with NaN values
* Output
  + Matrix with mean values for the nans

Distribution plot

* Matlab function of distribution plot: normplot
* Selfmade plot that:
  + Orders the data
  + Changes the colour if the values are outliers
  + Plots a line that represents the normal distribution

Input:

Class: matrix of the class data

Txtdata: textdata of the variables

n: class number for in the title

Outlier: string for ‘without outliers’ or ‘with outliers’

Subrows: row amount for subplot

Subcollumns: collumn amount for subplot

Subplotratio: defines the best ratio for the subplots

Input: number of variables

Output: row column ratio

Scatterfit: fits the data to slope

Input: data with columns are the different datapoint and rows the individuals

**LDA**

LDA: main LDA funtion

* Input:
  + Classes
  + the text data of the features (to sort them)
* Output:
  + LD1\_proj: mapped data to LD1 (1xn)
  + LD2\_proj: mapped data to LD2 (1xn)
  + Proj\_Vector: LD1 and LD2
  + V\_sorted: All the LDs sorted
  + D\_sorted: all the eigenvalues sorted
  + bartext\_LDA: the sorted text for the bar graphs of LD1 and LD2
  + bardata\_LDA: the sorted data for the bar graphs of LD1 and LD2
  + LDA\_summary: mapped data to each LD

The coloured outputs could be merged together, but because they get used for different purposes, you would need to extract them anyways, so I prefer to separate them

To plot 2 LDs to each other other than the first 2 you need a function that accepts 2 defined LDs and maps the data to those 2; the same calculation as the main LDA function but without the first part of calculation the LDs

LDAplot: the main LDA plot function with:

* Barplot of both the LDs:
  + Absolute magnitude of the contribution per feature with color coding for the sign
  + Cumulative value for first 3 and 10 features
* 2D scatter plot of mapped to LD1 vs LD2
  + Color coding per class
  + Every individual has their index show for their class\*

\*both classes have index going from 1 to 15 but because the values are mostly separated different indexing doesn’t really matter, moreover it does mostly matter on outliers, which are more isolated, so the individual can be deduced from combination of the class color and the index

LDvsLDplot: plots LD1, LD2, LD3 to each other in 3 2D plots and 1 3D plot

Input:

3 vectors that represent the 3 best variables

Class size

textdata(of the 3 variables)

LDAsubplot: plots a predefined LD to all the other LDs

Input:

2 projected LDs (LDn and LDm)

Class size

Index number of LDn

Index number of LDm

LDAPCAconv

* input
  + Proj\_Vector: the 2 project vectors from the LDA output each with dim: 1xn for n PC’s
  + Coef: the weighted data from the PCA
  + Txtdata the textdata to sort
* Output
  + bartext\_LDA: the sorted text
  + bardata\_LDA: the conversion of the the linear combination of the linear combination of features

LDA\_known: LDA plot for other LDs then LD1 and LD2

* Input:
  + Classes
  + the text data of the features (to sort them)
  + The 2 defined LDs
* Output:
  + LD1\_proj: mapped data to LD1 (1xn)
  + LD2\_proj: mapped data to LD2 (1xn)
  + bartext\_LDA: the sorted text for the bargraphs of LD1 and LD2
  + bardata\_LDA: the sorted data for the bargraphs of LD1 and LD2