Name: Joshua Damm

Student Number: s4036018

I performed Principal Components Regression (PCR) on the alexithymia dataset and investigated if I would keep the same number of components as in Principal Components Analysis (PCA). First, I preprocessed the data, separated the response variable (depression score, CESD) from the predictor variables (20 in total). Then I ran PCA on the scaled predictors and decided to keep 6 components based on the scree plot and Kaiser’s rule. Especially Kaiser’s rule suggests keeping 6 components for Eigenvalues >1 in the correlation matrix of the predictors.

After that, I ran PCR with 20 components on the training data using 10-fold cross validation. I plotted the mean squared prediction error (MSEP) and decided to keep 13 components as the MSEP remains rather flat including component 13 and rises sharply after that, suggesting a higher bias and/or variance of the model with more than 13 principal components (PC’s) in predicting the response variable. Then I fitted a PCR again on the training data with these 13 principal components and derived a mean squared error (MSE) of 106.1891 compared to 115.7976 using all 20 components. After deciding for the number of components I ran PCR with the selected components again on the full data. In total, 87.63% of the variance in the response could be explained by the model. When investigating the loadings of the principal components on the response variable, especially components 1, 12, and 13 seem be associated with the response, suggesting that not all the intermediate components between 1 and 12 are predictive of depression score and might capture other aspects related to alexithymia. Interestingly, when investigating the loadings principal components on the predictor variables, only x. describe (ability to describe one’s own emotions) loads negatively on the first principal component, and on component 12, which suggest a high association between this variable and depression score. X.going.on (knowing what is going on inside oneself) and x.right.words (findings right words for one’s feeling) are highly loaded on PC1 and moderately loaded on PC2, suggesting a high prediction capability on depression score. This makes sense as both variables might capture different nuanced aspects of an underlying latent concept regarding introspection and having a connection to one’s own emotions, which is both present in alexithymia and depression.