Most common mistakes in Coursework 1

Task 1

1.1 Linear regression

- No description of what the small difference of R2 on training and test means (no overfitting)
- Test set *not* standardised on the mean and standard deviation of the training set
- Model fitted without intercept

1.2 Ridge Regression

- No description of how bias and variance are affected by large penalty parameter
- No comparison to results from 1.1
- Model fitted without intercept

1.3 Relaxation of Lasso Regression

- No discussion on parameter-to-parameter differences between Ridge and Lasso taking into account potential sparsity.
- Incorrect implementation of Huber gradient
- R2 scores not computed

Task 2

2.1 kNN Classifiers

- No mention which value is optimal k
- Range to scan k was chosen too small

2.2 Random Forests

- Too few comments on code
- T-fold cross validation should be performed using the same subsets as in Task 2.1
- Grid search should at least include 3 values in each dimension
- Grid search must be implemented to scan hyperparameter together and not independently
- No confusion matrices plotted

2.3 Support Vector Machine

- True and false positive rates must be computed over range of hardness parameter in ROC curve
- Add more details to discussion of balanced vs unbalanced data (bias, variance)
- Range of values for hyperparameters in grid search too small

Task 3

3.1 Logistic regression and bagging

- Range of values for hyperparameters in grid search too small (take at least 3 values)

3.2 Kernelised SVM classifier

- RBF kernel not correctly implemented