2.1.1

- `n_estimators`: Sets the number of trees in the forest. Typically, the higher the number of trees, the better the performance of the random forest classifier. Setting the number of trees very high increases the time of training.
- `max_depth`: Pertains to the depth of the decision trees comprising the random forest. If we want to obtain pure leaves (i.e., that the samples at each node belong to the same class), the depth of the random forest is proportional to the difficulty of obtaining pure leaves. We could easily have an instance where the random forest would be very deep with a lot of nodes. As the depth increases the risk of overfitting increases. **Pruning** is the practice of setting a maximal depth of the different decision trees.
- `max_features`: Could not find anything in slides or book, but it is known that if, say, two features are highly correlated and both predicatively powerful, it might be the case that one of the features has a high estimate attached to it and the other one a low estimate attached. Thus, decreasing the number of bootstrapped features would increase the number of times that only one feature was featured in a given decision tree, resulting in more accurate estimates for both features. The downside of reducing the size of the feature bootstrap would be that we exclude features that have predictive power from the tree thus resulting in lower predictive power.
- `bootstrap`: With respect to the bootstrap in random forests, the number and size of the samples are important. Decreasing the size of samples imply lower probability for obtaining the same datapoints in different samples, i.e., increasing the randomness of the forest. This helps counteract overfitting and decreases the spread between train and test performance, however, at the cost of generally lower overall test performance.

2.1.2 if the size of the feature bootstrap equals the total number of features, every tree would have the same bootstrap of features (because the bootstrap of features is without replacement). Since the bootstrap of sample is done with replacement, we shouldn't be able to ensure that all trees are trained on the same data and thus are identical in this dimension.
2.1.3 The model performs relatively well with a 95 percent testing accuracy. The number of trees does included does not alter the conclusions much. The model performs much better than just guessing/baseline.

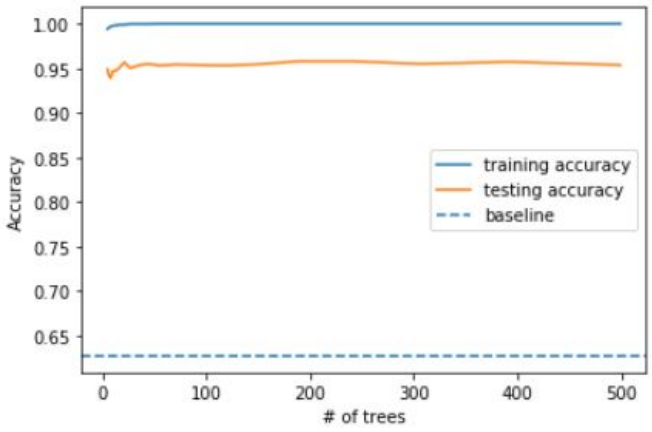
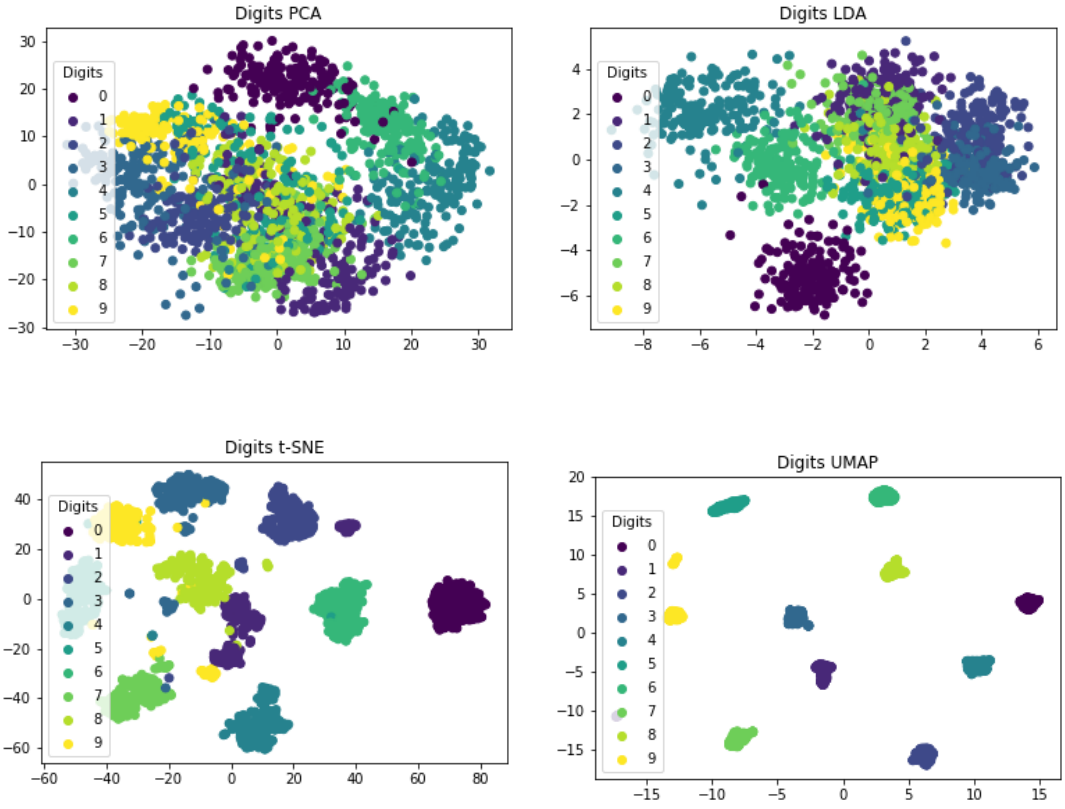


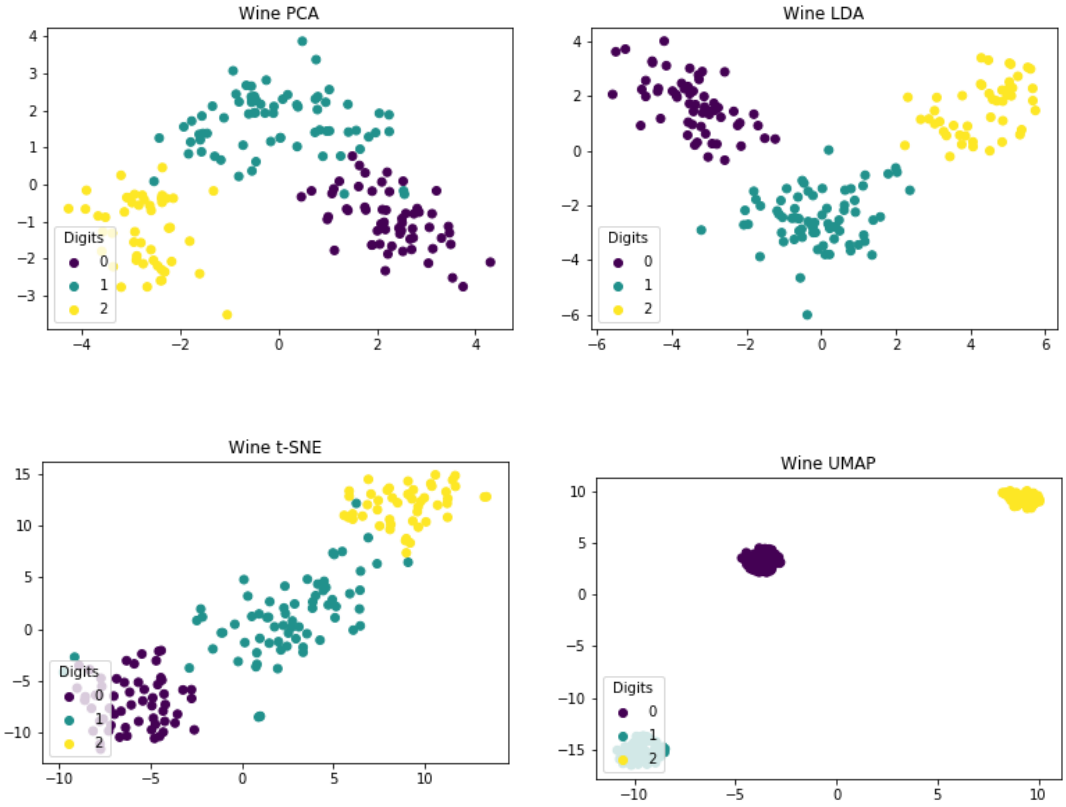
Fig. 1

3.1.2.1 As is clear from Fig. 1, PCA and LDA are identical to those of the slides. t-SNE and UMAP does differ, however, which is expected since they change every time the code is run. We see that PCA and LDA does not classify the images satisfactorily. t-SNE does a reasonable well job and UMAP comes out on top with impeccable classification. We need not standardisation because the data are the intensity of pixel and so already standardised.



Figur 1: Fig. 1

3.1.2.2 The wine datasets consists of data on 13 ingredients in three types of wine and the goal is from these ingredients to predict the type of wine. In contrast with 3.1.2.1. where the features were already directly comparable, here, we need to perform a standardisation such that the features are measured on the same scale. As is visible from Fig. 2, UMAP and LDA do the best job and PCA and t-SNE the worst.



Figur 2: Fig. 1

3.1.2.3. We see that PCA and LDA classifies poorly the digits dataset both does a good and perfect job for the wine dataset, respectively. t-SNE does fairly well for both datasets and UMAP classifies perfectly for both dataset. This shows there is a necessity for using different classifiers for different types of data.

4.1.2.1. Honest estimation involves splitting the training data into two parts: one part used for constructing the tree (including regularization such as cross-validation) and another part used for estimating treatment effects. The results is that the (asymptotic) properties of the estimator of the treatment estimator are the same as had the partitions in the tree been exogenously given. Hereby the bias attached adaptive estimation is eliminated. When the procedure of estimation is adaptive, (spurious) extreme values are likely to be placed into the same leaf as one another. This would lead to extreme values of means compared to an exogenously given split.
The drawback of honest estimation is the loss of data, i.e., the data used to determine partitions.

4.1.2.2. I think I described this above?