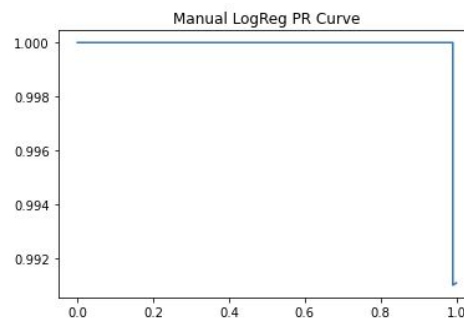


Logistic Regression (Manual)

Precision and recall computed with

`sklearn.metrics.precision_recall_curve`

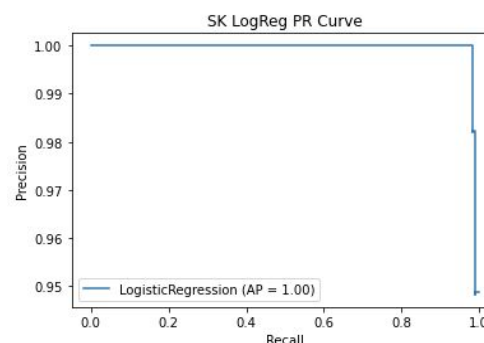
Trial	Test Error	Training Error
1	0.06735	0.00660

*Logistic Regression (sk-learn)*

Precision and recall computed with

`sklearn.metrics.plot_precision_recall_curve`

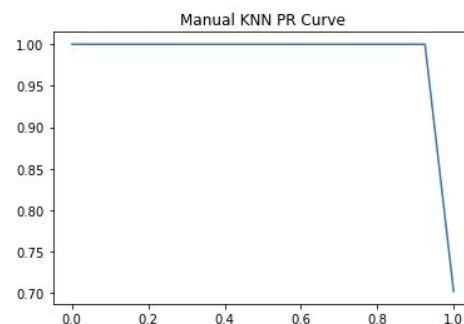
Trial	Test Error	Training Error
1	0.01169	0.01507

*KNN (Manual)*

Precision and recall computed with

`sklearn.metrics.precision_recall_curve`

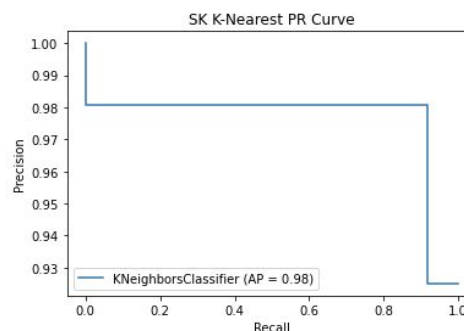
Trial	Test Error	Training Error
1	0.05263	0.03266

*KNN (sk-learn)*

Precision and recall computed with

`sklearn.metrics.plot_precision_recall_curve`

Trial	Test Error	Training Error
1	0.064327	0.02010

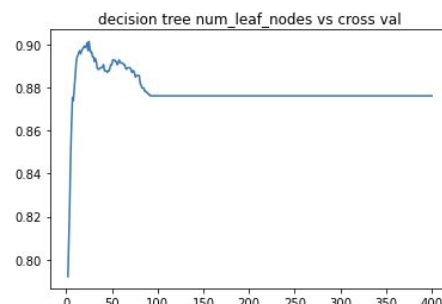
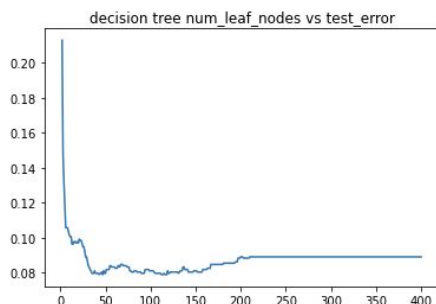


The sk-learn logistic regression classifier had the lowest test-error. The KNN algorithms were slower to train, with 30-dim data; however, the manual KNN out-performed the manual logistic regression classifier. KNN is slow and inaccurate when dealing with high-dimensional data, though this dataset was manageable and featured 96%+ test-scores for both KNN models.

Decision Tree Classifier

Evaluated decision trees with max-leaves across [2, 400].

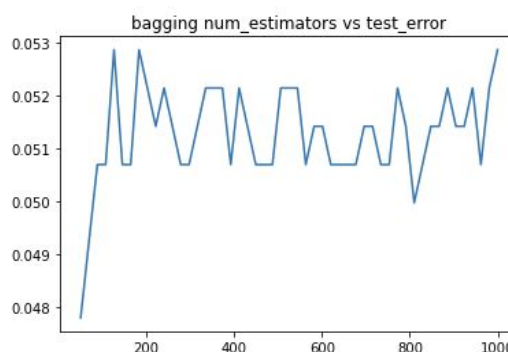
Trial	Test Error (best)	Max Leaves
1	0.0651	110



50 Bagging Classifier

Evaluated bagging classifiers with [50, 2500] estimators at intervals of 7.

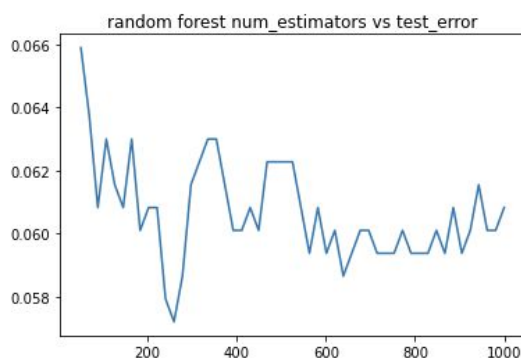
Trial	Test Error (best)	Estimators	Elapsed (s)
1	0.0447791	50	728



50 Random Forests

Evaluated random forest classifiers with [50, 2500] estimators at intervals of 7.

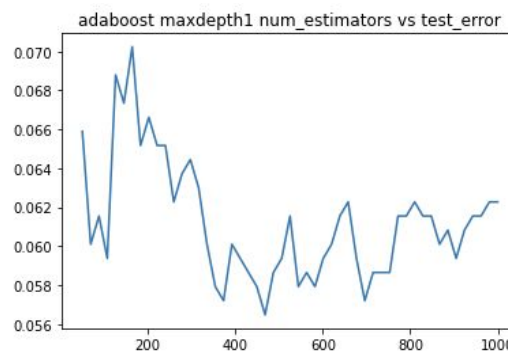
Trial	Test Error (best)	Estimators	Elapsed (s)
1	0.05720	107	700



50 AdaBoost Classifiers (max-depth=1)

Evaluated random forest classifiers with [50, 2500] estimators at intervals of 7. Base-classifier is a Decision Tree with `max_depth=1`.

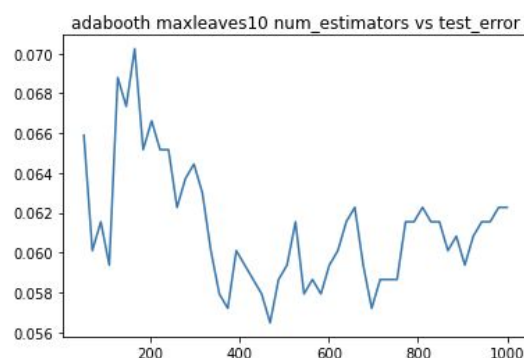
Trial	Test Error (best)	Estimators	Elapsed (s)
1	0.05638	259	107



50 AdaBoost Classifiers (max-leaf-nodes=10)

Evaluated every 7th random forest classifier with `num_estimator` in [50, 2500]. The base-classifier is a decision tree with `max_leaf_nodes=10`.

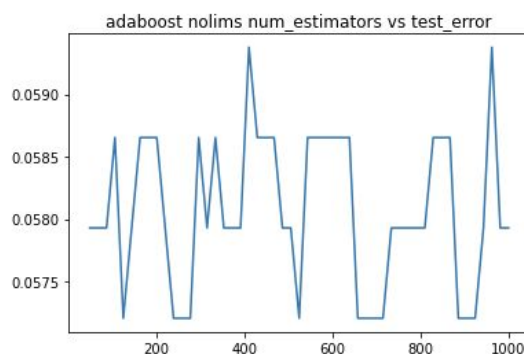
Trial	Test Error (best)	Estimators	Elapsed (s)
1	0.05641	468	388



50 AdaBoost Classifiers (no limitations)

Evaluated every 7th AdaBoost classifier with `num_estimators` in [50, 2500]. The base-classifier is a decision tree with no limitations (i.e. `max_leaf_nodes=None`, `max_depth=None`).

Trial	Test Error (best)	Estimators	Elapsed (s)
1	0.057205	126	441



The test-errors of all models were around 5%. Looking at the various ensemble methods, performance was typically best when the number of estimators was low (in the 100-300 range). Looking at the `num_estimators` vs `test_error` plots, the error fluctuated significantly across the entire domain.