

```

import warnings
warnings.filterwarnings('ignore')

import numpy as np
import pandas as pd
from datetime import datetime
from scipy.io import loadmat
from sklearn.decomposition import PCA
from sklearn.model_selection import (train_test_split, GridSearchCV)
from sklearn.metrics import (confusion_matrix, roc_curve, roc_auc_score)
from sklearn.cross_validation import KFold
from sklearn.ensemble import (RandomForestClassifier, AdaBoostClassifier,
                               GradientBoostingClassifier)
from xgboost.sklearn import XGBClassifier

%matplotlib inline
import matplotlib.pyplot as plt
import seaborn as sns

```

/home/cab/.local/lib/python3.5/site-packages/sklearn/cross_validation.py:41: DeprecationWarning: This module was deprecated in version 0.18 in favor of the model_selection module into which all the refactored classes and functions are moved. Also note that the interface of the new CV iterators are different from that of this module. This module will be removed in 0.20.

"This module will be removed in 0.20.", DeprecationWarning)

```
start = datetime.now().isoformat()
```

```

def biplot(setting, pca, x_reduced, y_train):
    pca_score = x_reduced[:,0:2]
    pca_coeff = pca.components_[0:2, :].T

    # Plot setup
    limit = { 'mnist': 0.5, 'mnist_zoomed': 0.15, 'sheep': 0.75 }
    f = plt.figure(figsize=(12, 12))
    plt.xlim(-limit[setting], limit[setting])
    plt.ylim(-limit[setting], limit[setting])
    ax = f.gca()
    radius = { 'mnist': 0.01, 'mnist_zoomed': 0.01, 'sheep': 0.05 }
    for i in range(5, 15, 1):
        rcircle = plt.Circle( (0,0), i * radius[setting], edgecolor="black", facecolor="
none", alpha=0.8)
        ax.add_artist(rcircle)
    plt.xlabel("PC1", fontsize=16)
    plt.ylabel("PC2", fontsize=16)
    plt.grid()

    # Draw the data points.
    scatter_alpha = { 'mnist': 0.4, 'mnist_zoomed': 0.3, 'sheep': 0.4 }
    xs = pca_score[:,0]
    ys = pca_score[:,1]
    scalex = 1.0/(xs.max() - xs.min())
    scaley = 1.0/(ys.max() - ys.min())
    plt.scatter(xs * scalex, ys * scaley, c=y_train, s=25, alpha=scatter_alpha[setting])

    # Draw the variable vectors and names.
    var_fontsize = { 'mnist': 8, 'mnist_zoomed': 15, 'sheep': 20 }
    for i in range(p):
        plt.arrow(0, 0, pca_coeff[i,0], pca_coeff[i,1],
                  color = 'black', alpha = 0.5, width=0.0001)
        plt.text(pca_coeff[i,0]* 1.05, pca_coeff[i,1] * 1.05, str(i),
                  color = 'navy', ha = 'center', va = 'center', fontsize=var_fontsize[set
ting])

```

```

plt.show()

def gridsearchcv_fit(clf, params, x_train, y_train):
    gs = GridSearchCV(clf, params, cv=10, scoring='roc_auc', n_jobs=-1, verbose=1)
    gs.fit(x_train, y_train)
    print("Search found the best params: {} \n ROC AUC Train Score: {}".format(
        gs.best_params_, gs.best_score_))
    return gs.best_estimator_

def precision_recall(cm):
    fp, fn, tp = cm[0][1], cm[1][0], cm[1][1]
    precision = tp / (tp + fp)
    recall = tp / (tp + fn)
    return precision * 100, recall * 100

def plot_roc(y_test, y_score):
    fpr, tpr, th = roc_curve(y_true=y_test, y_score=y_score)
    plt.plot(fpr, tpr, linewidth=2)
    plt.plot([0, 1], [0, 1], 'k--')
    plt.axis([0, 1, 0, 1])
    plt.xlabel('False Positive Rate')
    plt.ylabel('True Positive Rate')

def get_oob(clf, x_train, y_train, x_test, random_state):
    n_folds = 10
    n_train = x_train.shape[0]
    n_test = x_test.shape[0]
    oob_train = np.zeros((n_train,))
    oob_test = np.zeros((n_test,))
    oob_test_skf = np.empty((n_folds, n_test))

    kf = KFold(n_train, n_folds=n_folds, shuffle=True, random_state=random_state)

    for i, (train_index, test_index) in enumerate(kf):
        x_tr = x_train[train_index]
        y_tr = y_train[train_index]
        x_te = x_train[test_index]

        clf.fit(x_tr, y_tr)

        oob_train[test_index] = clf.predict(x_te)
        oob_test_skf[i, :] = clf.predict(x_test)

    oob_test[:] = np.around(oob_test_skf.mean(axis=0))
    return oob_train.reshape(-1, 1), oob_test.reshape(-1, 1)

def barchart_double(min_yylim, max_yylim, n_groups, list1, label1, list2, label2):
    fig, ax = plt.subplots(figsize=(15,5))
    ax.set(yylim=[min_yylim, max_yylim])
    #plt.figure(figsize=(15,5))
    index = np.arange(n_groups)
    bar_width = 0.35
    opacity = 0.8

    rects1 = plt.bar(index, list1, bar_width,
                      alpha=opacity,
                      color='b',
                      label=label1)

    rects2 = plt.bar(index + bar_width, list2, bar_width,
                      alpha=opacity,
                      color='g',
                      label=label2)

    plt.xticks(index + bar_width, ('Random Forest', 'AdaBoost', 'XGBoost', 'Majority Vo

```

```
te', 'Meta-Forest', 'Meta-XGBoost'))
plt.legend()

plt.tight_layout()
plt.show()
```

Overview

- The main aim of this analysis is to compare the binary classification performance of popular ensembling algorithms and methods: Random Forest, AdaBoost, XGBoost; and two stacking methods: Majority Vote and Meta Learning (Random Forest and XGBoost)
- For a broader picture, the comparison is performed on two datasets.
- The difference between them is that the MNIST response is heavily skewed, where one class vastly outnumbers the other.
- A secondary aim is to observe if any algorithm may perform well on precision or recall while being outperformed by another algorithm overall. It is expected that this may manifest more noticeably on the skewed MNIST data.
- The overall measurement used is the ROC AUC score for a more relevant comparison to the Sheep data where precision/recall is expected to be less relevant.

MNIST Dataset Preparation

```
mnist = loadmat('mnist-original.mat')
mnist.keys()
```

```
dict_keys(['data', '__globals__', '__version__', '__header__', 'mldata_descr_ordering', 'label'])
```

The MNIST data set is a large database of images of handwritten digits. Each instance is made of 784 dimensions which are a 1 dimensional reshape of a 28 x 28 pixel image. It is assumed that many features would be highly correlated due to the visual nature of the images. They are colored in grayscale, and large areas (particularly the outer areas) are plain white. The shapes that make up digits are simple, especially considering that each digit is presented in isolation as opposed to surrounded by other letters. Further, this problem will be further simplified by reducing the multi class label to binary and the simplest digit, 1, is chosen as the positive class.

```
data = mnist['data'].T
label = (mnist['label'] == 1).astype(int).T
n, p = data.shape
classes, counts = np.unique(label, return_counts=True)
print("n = {}\np = {}\nclass, count:\n{}".format(n, p, list(zip(classes, counts))))
```

```
n = 70000
p = 784
class, count:
[(0, 62123), (1, 7877)]
```

As both datasets differ in class counts, they will both be converted to a binary classification problem. For this dataset, the label value of 1 denotes the handwritten image is classified as the digit 1, and 0 denotes all digits (0 to 9) excluding 1.

```
plt.pie(counts, labels=classes)
plt.show()
```



Figure 1: Proportion of classes in the data is heavily skewed which may influence or explain performance differences of the learning methods between this dataset and the sheep dataset (which is later found to be relatively balanced).

Split data into train/test

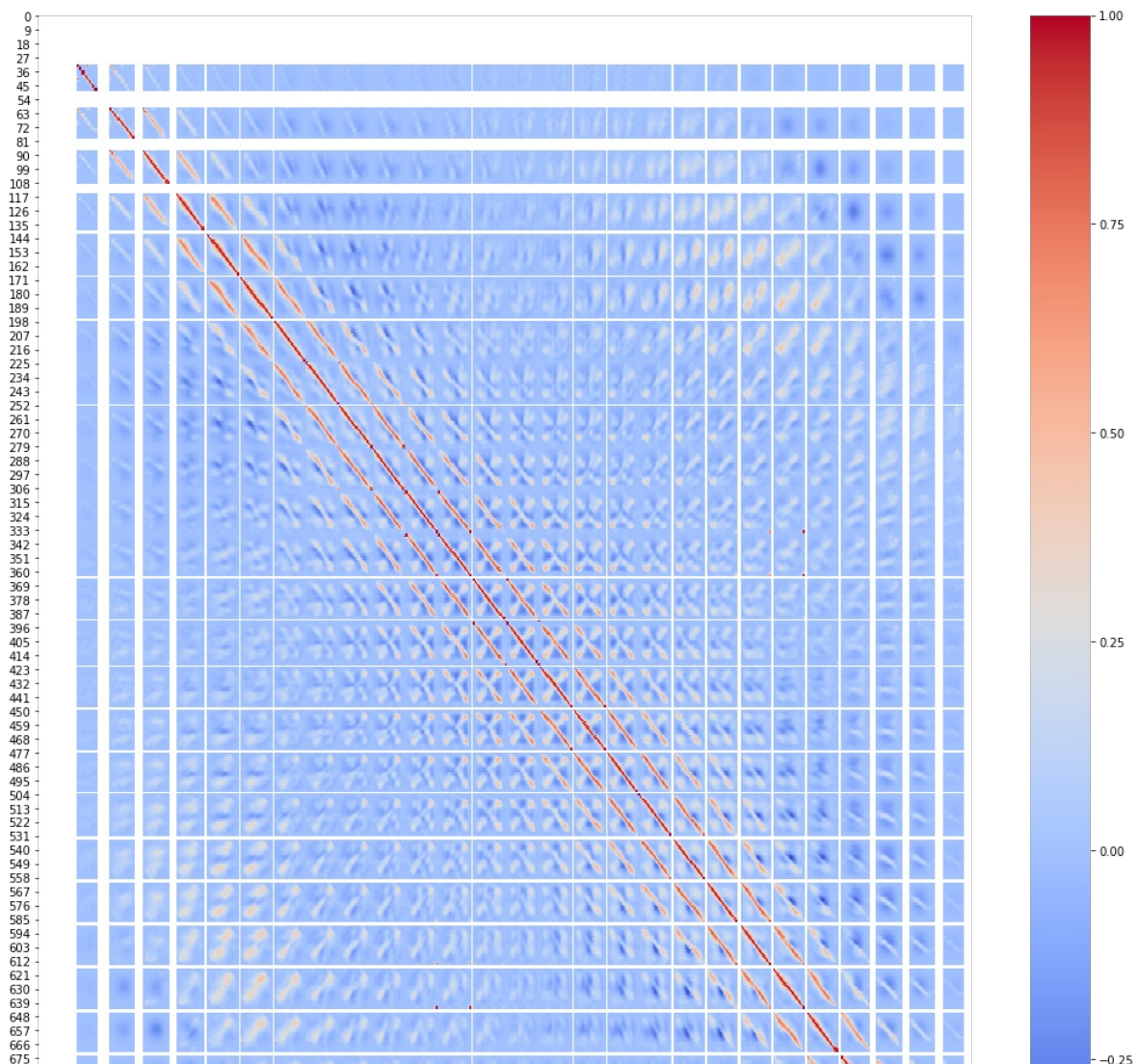
```
seed = 0
x_train_split, x_test_split, y_train, y_test = train_test_split(data, label, train_size
=4000, random_state=seed)
y_train, y_test = y_train.ravel(), y_test.ravel()
print("n_train, n_test: {0}, {1}".format(len(y_train), len(y_test)))
```

```
n_train, n_test: 4000, 66000
```

The traditional 80%/20% train/split ratio would push training times too long considering the objectives of this analysis. For both datasets, 4000 instances are allocated for training, and the remaining for testing.

Correlation Heatmap

```
f, ax = plt.subplots(figsize=(18, 18))
corr = pd.DataFrame(x_train_split).corr()
hm = sns.heatmap(round(corr,2), annot=False, ax=ax, cmap="coolwarm",fmt='.2f')
f.subplots_adjust(top=0.93)
```



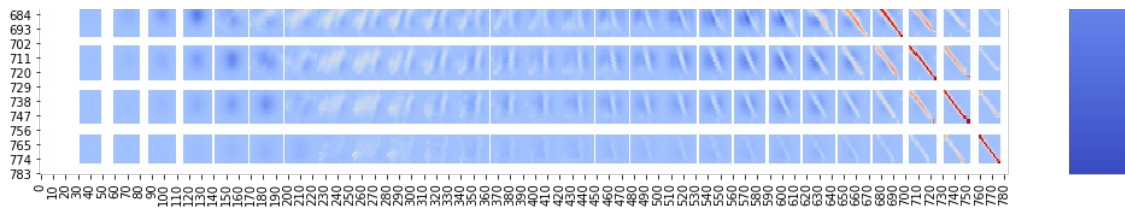


Figure 2: Correlation heatmap of the features reveals an interesting, distinguishable pattern. The parallel red lines following the main diagonal appear to be approximately 24 features apart. Other colour patterns follow a similar distance between similar colours. Considering how the features are a 1 dimensional reshape of a 24 x 24 pixel image, the patterns support the assumption that adjacent pixels are heavily correlated.

Principle Component Analysis

```
pca_init = PCA(n_components=0.95)
x_reduced1 = pca_init.fit_transform(x_train_split)
print("Number of Dimensions:\n{n:3d} - original\n{n:3d} - reduced ({} dimensions
eliminated) "
      .format(p, x_reduced1.shape[1], p - x_reduced1.shape[1]))
```

```
Number of Dimensions:
784 - original
147 - reduced (637 dimensions eliminated)
```

The PCA is initially run to maximally reduce the number of dimensions while retaining information that explains 95% of the variance. This transformation has eliminated a large majority dimensions, adding to evidence of high correlation and redundant information.

```
plt.plot(np.cumsum(pca_init.explained_variance_ratio_))
plt.xlabel("Number of Dimensions")
plt.ylabel("Proportion of Variance Explained")
plt.show()
```

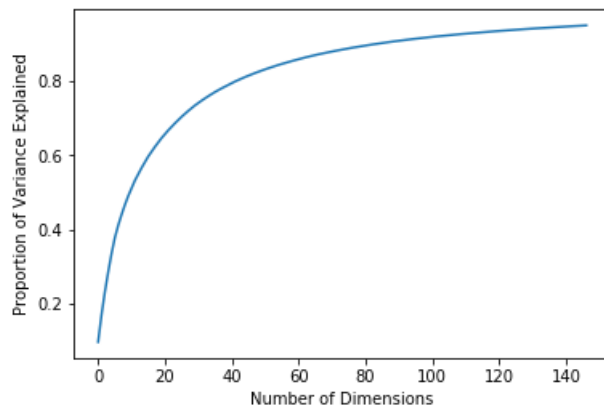


Figure 3: The proportion of variance explained plots a smooth curve, i.e. there appears to be no obvious "elbow". A dimension value within the range of 25 and 50 could be argued as a range that appears most "elbow-like". The proportion value of 80% lies within this range and is thus chosen to run another PCA to further reduce dimensions and achieve faster training speeds. The plot shows that 15 percentage points of explained variance can be sacrificed for discarding approximately 100 dimensions.

```
pca_final = PCA(n_components=0.8)
x_reduced2 = pca_final.fit_transform(x_train_split)
print("Number of Dimensions:\n{n:3d} - original\n{n:3d} - reduced ({} dimensions
eliminated) "
      .format(p, x_reduced2.shape[1], p - x_reduced2.shape[1]))
```

```
Number of Dimensions:
784 - original
43 - reduced (741 dimensions eliminated)
```


A further 100 dimensions are eliminated for a sacrifice of 15% percentage points. Given the context of this analysis and the hardware running these algorithms, it is considered a worthwhile tradeoff to train on this transformation rather than the original 784 dimension data.

PCA Biplot

```
pclevr, pc2evr = np.multiply(pca_final.explained_variance_ratio_[:2], np.full(2, 100))
print("Proportion of Variance Explained:\n{:.2f}% - PC1\n{:.2f}% - PC2\n{:.2f}% - PC1\nand PC2"
      .format(pclevr, pc2evr, pclevr + pc2evr))
```

```
Proportion of Variance Explained:
9.61% - PC1
7.18% - PC2
16.79% - PC1 and PC2
```

The first 2 components that will be plotted contribute to 16.79% of the variance. Although this is a significant amount in relation to the high dimensions, there will still be a large majority of variance missing from the following plots.

```
biplot('mnist', pca_final, x_reduced2, y_train)
```

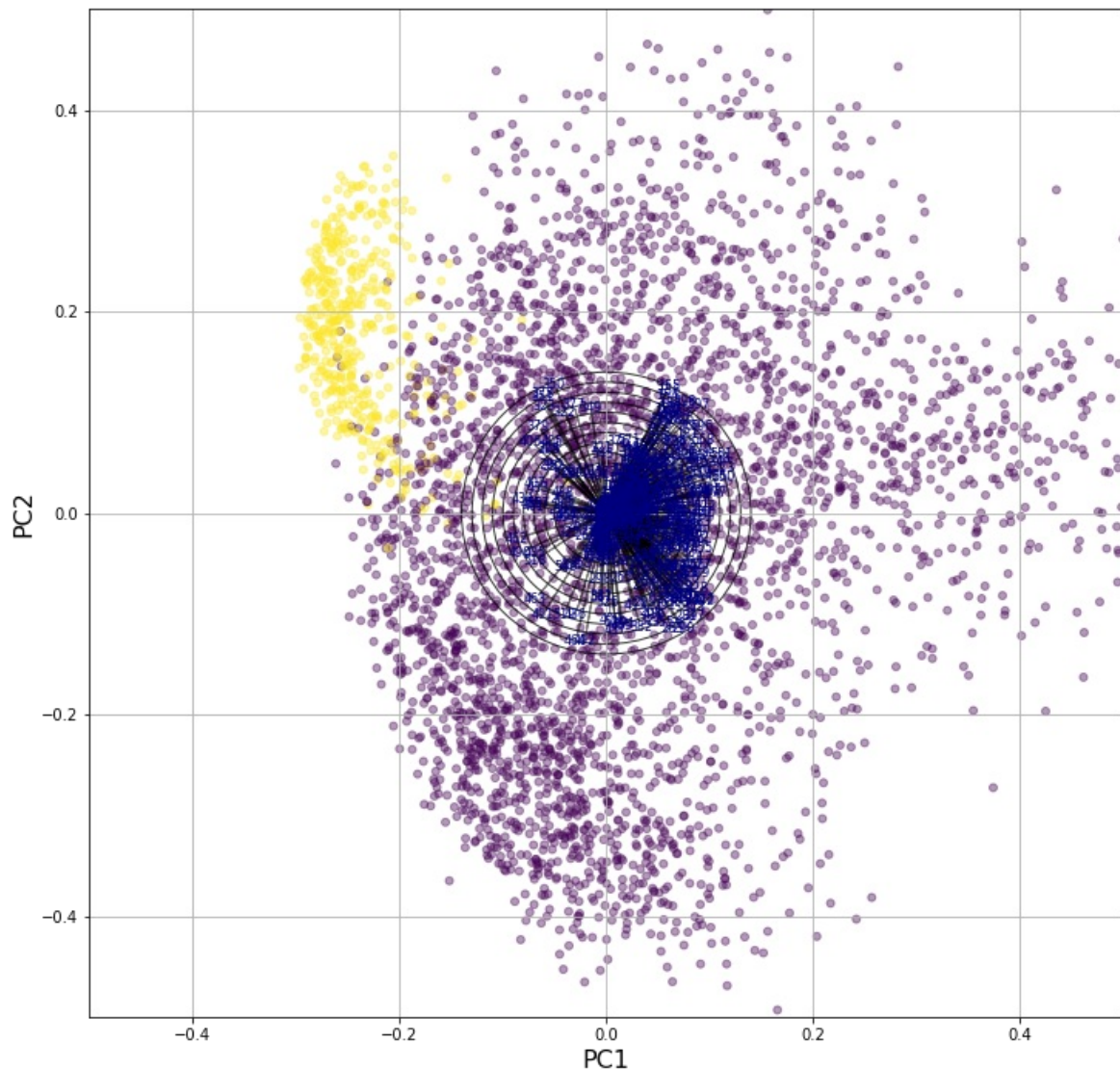


Figure 3: Biplot of the first two principle components. The magnitudes are found to be relatively small to the distances between opposing outer data points. A zoomed in version is given in Figure 3.

```
biplot('mnist_zoomed', pca_final, x_reduced2, y_train)
```

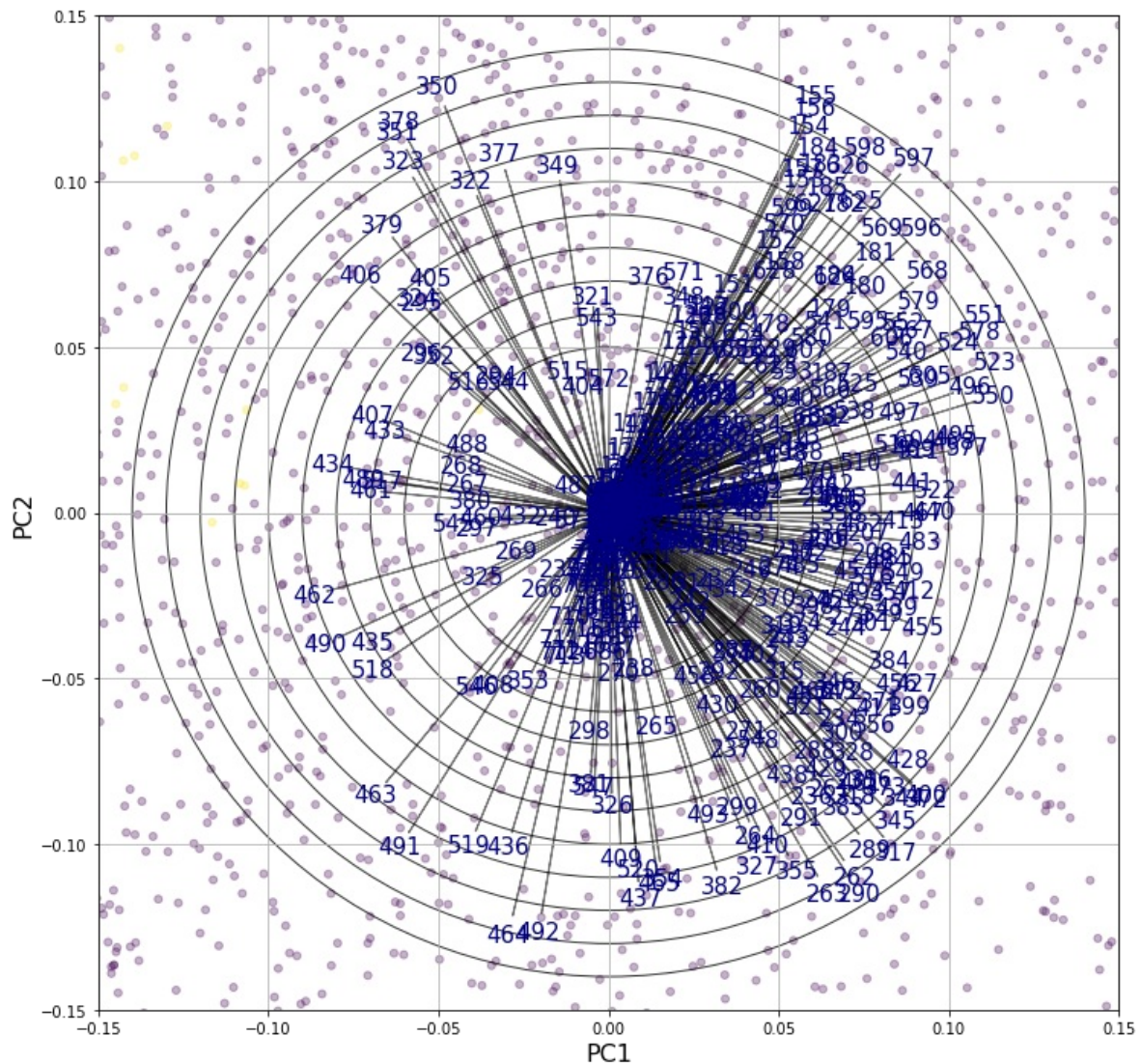


Figure 4: Zoomed version of Figure 2. Out of 784 variables, it appears that a relatively small proportion (50-100) of them outgrow all others in magnitude. Further, there are noticeable, concentrations of direction; towards the right and slightly towards the top left.

The accumulated evidence of high correlation reflects the dataset's main characteristics: each image surrounds the handwritten digit in a greater area of white space, and adjacent pixels are likely to be similar in intensity.

Model Training

- k = 10 folds are used for cross validation to search for the best params.
- The ROC AUC method of scoring is chosen to determine the model with the best params.
- Both the ROC AUC and precision/recall methods of scoring are chosen to record and compare all final models between both datasets.
- Where applicable, the maximum number of available threads for training computation is used.
- For this dataset, the previous pca transformation is applied to reduce the dataset's dimensionality for faster training times.

```
n_jobs = -1
auc_mnist = {}
pr_mnist = {}
x_train = pca_final.transform(x_train_split)
x_test = pca_final.transform(x_test_split)
x_train.shape, x_test.shape
```

```
((4000, 43), (66000, 43))
```

Random Forest

```
p_reduced = x_test.shape[1]
rf_params = [
    {'n_estimators': [ 1500, 1750, 2000, 2250 ], 'max_features': [ "sqrt",
int(p_reduced / 3) ]}
]
```

- `n_estimators`: The number of trees.
- `max_features`: `m` = the size of the subset of features considered for each tree. "sqrt" denotes the $m = \sqrt{p}$ setting that is mostly recommended for classification, while the next setting, $m = p / 3$, is recommended for regression.

```
rf_init = RandomForestClassifier(random_state=seed, n_jobs=n_jobs, verbose=1)
rf = gridsearchcv_fit(rf_init, rf_params, x_train, y_train)
```

Fitting 10 folds for each of 8 candidates, totalling 80 fits

```
[Parallel(n_jobs=-1)]: Done 34 tasks      | elapsed:    0.2s
[Parallel(n_jobs=-1)]: Done 34 tasks      | elapsed:    0.2s
[Parallel(n_jobs=-1)]: Done 34 tasks      | elapsed:    0.3s
[Parallel(n_jobs=-1)]: Done 34 tasks      | elapsed:    0.3s
[Parallel(n_jobs=-1)]: Done 34 tasks      | elapsed:    0.3s
[Parallel(n_jobs=-1)]: Done 34 tasks      | elapsed:    0.3s
[Parallel(n_jobs=-1)]: Done 34 tasks      | elapsed:    0.4s
[Parallel(n_jobs=-1)]: Done 34 tasks      | elapsed:    0.4s
[Parallel(n_jobs=-1)]: Done 184 tasks     | elapsed:    1.7s
[Parallel(n_jobs=-1)]: Done 184 tasks     | elapsed:    1.8s
[Parallel(n_jobs=-1)]: Done 184 tasks     | elapsed:    1.8s
[Parallel(n_jobs=-1)]: Done 184 tasks     | elapsed:    1.8s
[Parallel(n_jobs=-1)]: Done 184 tasks     | elapsed:    1.8s
[Parallel(n_jobs=-1)]: Done 184 tasks     | elapsed:    2.1s
[Parallel(n_jobs=-1)]: Done 184 tasks     | elapsed:    2.4s
[Parallel(n_jobs=-1)]: Done 434 tasks     | elapsed:    4.2s
[Parallel(n_jobs=-1)]: Done 434 tasks     | elapsed:    4.3s
[Parallel(n_jobs=-1)]: Done 434 tasks     | elapsed:    4.3s
[Parallel(n_jobs=-1)]: Done 434 tasks     | elapsed:    4.3s
[Parallel(n_jobs=-1)]: Done 434 tasks     | elapsed:    4.2s
[Parallel(n_jobs=-1)]: Done 434 tasks     | elapsed:    4.3s
[Parallel(n_jobs=-1)]: Done 434 tasks     | elapsed:    4.8s
[Parallel(n_jobs=-1)]: Done 434 tasks     | elapsed:    5.4s
[Parallel(n_jobs=-1)]: Done 784 tasks     | elapsed:    7.4s
[Parallel(n_jobs=-1)]: Done 784 tasks     | elapsed:    7.6s
[Parallel(n_jobs=-1)]: Done 784 tasks     | elapsed:    7.7s
[Parallel(n_jobs=-1)]: Done 784 tasks     | elapsed:    7.9s
[Parallel(n_jobs=-1)]: Done 784 tasks     | elapsed:    8.0s
[Parallel(n_jobs=-1)]: Done 784 tasks     | elapsed:    7.9s
[Parallel(n_jobs=-1)]: Done 784 tasks     | elapsed:    8.1s
[Parallel(n_jobs=-1)]: Done 784 tasks     | elapsed:    9.4s
[Parallel(n_jobs=-1)]: Done 1234 tasks    | elapsed:   12.1s
[Parallel(n_jobs=-1)]: Done 1234 tasks    | elapsed:   12.2s
[Parallel(n_jobs=-1)]: Done 1234 tasks    | elapsed:   12.3s
[Parallel(n_jobs=-1)]: Done 1234 tasks    | elapsed:   12.4s
[Parallel(n_jobs=-1)]: Done 1234 tasks    | elapsed:   12.5s
[Parallel(n_jobs=-1)]: Done 1234 tasks    | elapsed:   12.8s
[Parallel(n_jobs=-1)]: Done 1234 tasks    | elapsed:   12.8s
[Parallel(n_jobs=-1)]: Done 1234 tasks    | elapsed:   14.0s
[Parallel(n_jobs=-1)]: Done 1500 out of 1500 | elapsed:   14.9s finished
[Parallel(n_jobs=-1)]: Done 1500 out of 1500 | elapsed:   14.9s finished
[Parallel(n_jobs=-1)]: Done 1500 out of 1500 | elapsed:   14.9s finished
[Parallel(n_jobs=-1)]: Done 1500 out of 1500 | elapsed:   14.9s finished
[Parallel(n_jobs=-1)]: Done 1500 out of 1500 | elapsed:   15.2s finished
[Parallel(n_jobs=8)]: Done 34 tasks      | elapsed:    0.1s
[Parallel(n_jobs=8)]: Done 34 tasks      | elapsed:    0.1s
[Parallel(n_jobs=8)]: Done 34 tasks      | elapsed:    0.0s
[Parallel(n_jobs=8)]: Done 34 tasks      | elapsed:    0.0s
[Parallel(n_jobs=-1)]: Done 1500 out of 1500 | elapsed:   15.2s finished
[Parallel(n_jobs=-1)]: Done 1500 out of 1500 | elapsed:   15.3s finished
[Parallel(n_jobs=8)]: Done 184 tasks     | elapsed:    0.2s
[Parallel(n_jobs=8)]: Done 34 tasks      | elapsed:    0.0s
[Parallel(n_jobs=8)]: Done 184 tasks     | elapsed:    0.1s
[Parallel(n_jobs=8)]: Done 184 tasks     | elapsed:    0.2s
[Parallel(n_jobs=8)]: Done 184 tasks     | elapsed:    0.1s
[Parallel(n_jobs=8)]: Done 184 tasks     | elapsed:    0.1s
```


[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

```

[Parallel(n_jobs=8)]: Done 784 tasks      | elapsed:    0.4s
[Parallel(n_jobs=8)]: Done 784 tasks      | elapsed:    0.4s
[Parallel(n_jobs=8)]: Done 784 tasks      | elapsed:    0.4s
[Parallel(n_jobs=8)]: Done 784 tasks      | elapsed:    0.4s
[Parallel(n_jobs=8)]: Done 434 tasks      | elapsed:    0.2s
[Parallel(n_jobs=8)]: Done 784 tasks      | elapsed:    0.4s
[Parallel(n_jobs=8)]: Done 784 tasks      | elapsed:    0.4s
[Parallel(n_jobs=8)]: Done 784 tasks      | elapsed:    0.4s
[Parallel(n_jobs=8)]: Done 784 tasks      | elapsed:    0.4s
[Parallel(n_jobs=8)]: Done 1234 tasks     | elapsed:    0.7s
[Parallel(n_jobs=8)]: Done 1234 tasks     | elapsed:    0.7s
[Parallel(n_jobs=8)]: Done 1234 tasks     | elapsed:    0.7s
[Parallel(n_jobs=8)]: Done 1234 tasks     | elapsed:    0.7s
[Parallel(n_jobs=8)]: Done 1234 tasks     | elapsed:    0.7s
[Parallel(n_jobs=8)]: Done 1234 tasks     | elapsed:    0.7s
[Parallel(n_jobs=8)]: Done 1234 tasks     | elapsed:    0.7s
[Parallel(n_jobs=8)]: Done 1784 tasks     | elapsed:    1.0s
[Parallel(n_jobs=8)]: Done 1784 tasks     | elapsed:    1.0s
[Parallel(n_jobs=8)]: Done 1784 tasks     | elapsed:    1.0s
[Parallel(n_jobs=8)]: Done 1784 tasks     | elapsed:    1.0s
[Parallel(n_jobs=8)]: Done 1784 tasks     | elapsed:    1.0s
[Parallel(n_jobs=8)]: Done 1784 tasks     | elapsed:    1.0s
[Parallel(n_jobs=8)]: Done 1784 tasks     | elapsed:    1.0s
[Parallel(n_jobs=8)]: Done 2250 out of 2250 | elapsed:    1.2s finished
[Parallel(n_jobs=8)]: Done 2250 out of 2250 | elapsed:    1.2s finished
[Parallel(n_jobs=8)]: Done 2250 out of 2250 | elapsed:    1.2s finished
[Parallel(n_jobs=8)]: Done 2250 out of 2250 | elapsed:    1.3s finished
[Parallel(n_jobs=8)]: Done 2250 out of 2250 | elapsed:    1.2s finished
[Parallel(n_jobs=8)]: Done 2250 out of 2250 | elapsed:    1.2s finished
[Parallel(n_jobs=8)]: Done 2250 out of 2250 | elapsed:    1.1s finished
[Parallel(n_jobs=-1)]: Done 80 out of 80 | elapsed: 6.1min finished
[Parallel(n_jobs=-1)]: Done 34 tasks      | elapsed:    0.1s
[Parallel(n_jobs=-1)]: Done 184 tasks     | elapsed:    0.3s
[Parallel(n_jobs=-1)]: Done 434 tasks     | elapsed:    0.7s
[Parallel(n_jobs=-1)]: Done 784 tasks     | elapsed:    1.3s
[Parallel(n_jobs=-1)]: Done 1234 tasks    | elapsed:    2.0s
[Parallel(n_jobs=-1)]: Done 1784 tasks    | elapsed:    2.9s

```

Search found the best params: {'max_features': 'sqrt', 'n_estimators': 2000}
ROC AUC Train Score: 0.998492175760427

```

[Parallel(n_jobs=-1)]: Done 2000 out of 2000 | elapsed:    3.3s finished

```

rf

```

RandomForestClassifier(bootstrap=True, class_weight=None, criterion='gini',
                        max_depth=None, max_features='sqrt', max_leaf_nodes=None,
                        min_impurity_decrease=0.0, min_impurity_split=None,
                        min_samples_leaf=1, min_samples_split=2,
                        min_weight_fraction_leaf=0.0, n_estimators=2000, n_jobs=-1,
                        oob_score=False, random_state=0, verbose=1, warm_start=False)

```

The random forest model according to Breiman involves bootstrapping which is enabled here as default. In this model, the common criterion of the Gini Index is used to determine the split at each tree node. For classification, the recommended setting of "sqrt" for classification is confirmed to be optimal here. With this setting, 2000 trees were found to produce the best ROC before overfitting at 2250 trees.

Feature Importances

```

f = plt.figure(figsize=(15,5))
bar_x_axis = range(len(rf.feature_importances_))
plt.bar(bar_x_axis, rf.feature_importances_, width=0.9, color="g",
        tick_label=bar_x_axis)
plt.tick_params(labelsize=10)
plt.show()

```

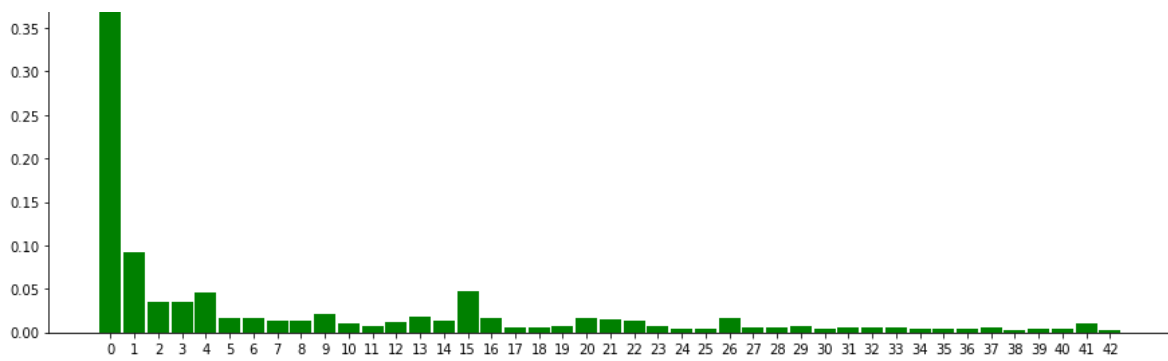


Figure 5: Bar chart of the Random Forest's feature importances. Naturally, the Random Forest finds the first few features (i.e. the most useful principal components) to be the most important.

Prediction Performance

```
rf_y_pred = rf.predict(x_test)
rf_cm = confusion_matrix(y_true=y_test, y_pred=rf_y_pred)
rf_cm
```

```
[Parallel(n_jobs=8)]: Done 34 tasks      | elapsed: 0.0s
[Parallel(n_jobs=8)]: Done 184 tasks     | elapsed: 0.2s
[Parallel(n_jobs=8)]: Done 434 tasks     | elapsed: 0.4s
[Parallel(n_jobs=8)]: Done 784 tasks     | elapsed: 0.7s
[Parallel(n_jobs=8)]: Done 1234 tasks    | elapsed: 1.1s
[Parallel(n_jobs=8)]: Done 1784 tasks    | elapsed: 1.6s
[Parallel(n_jobs=8)]: Done 2000 out of 2000 | elapsed: 1.8s finished
```

```
array([[58466, 96],
       [ 481, 6957]])
```

The Random Forest model has predicted 58466 true negatives and 6957 true positives.

```
pr_mnist['Random Forest'] = precision_recall(rf_cm)
pr_mnist
```

```
{'Random Forest': (98.6388770735857, 93.53320785157301)}
```

The Random Forest model scored a positive prediction accuracy of 98.64% (precision).
The rate of positive instances that are correctly detected is 93.533% (recall/sensitivity).
The model is found to be more precise than sensitive.

```
rf_y_score = rf.predict_proba(x_test)[: , 1]
plot_roc(y_test, rf_y_score)
auc_mnist['Random Forest'] = roc_auc_score(y_true=y_test, y_score=rf_y_score)
print("ROC AUC Test Score: {}".format(auc_mnist['Random Forest']))
plt.show()
```

```
[Parallel(n_jobs=8)]: Done 34 tasks      | elapsed: 0.0s
[Parallel(n_jobs=8)]: Done 184 tasks     | elapsed: 0.2s
[Parallel(n_jobs=8)]: Done 434 tasks     | elapsed: 0.4s
[Parallel(n_jobs=8)]: Done 784 tasks     | elapsed: 0.7s
[Parallel(n_jobs=8)]: Done 1234 tasks    | elapsed: 1.2s
[Parallel(n_jobs=8)]: Done 1784 tasks    | elapsed: 1.7s
[Parallel(n_jobs=8)]: Done 2000 out of 2000 | elapsed: 1.9s finished
```

```
ROC AUC Test Score: 0.9981555217541017
```



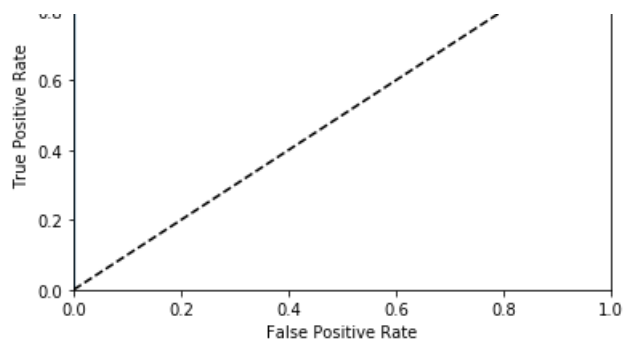


Figure 6: The Random Forest's ROC Curve on the test data.

AdaBoost

```
ab_params = [
    {'n_estimators': [ 800, 1000, 1200, 1400 ], 'learning_rate': [ 0.005, 0.01, 0.05,
0.1 ]}
]
```

- **n_estimators**: The number of trees.
- **learning_rate**: Applied to the weight of each tree, shrinking its contribution to the model. Lower rates are expected to better detect subtle features while requiring more trees to surpass underfitting. Higher rates risk premature overfitting. Compared to the Random Forest grid search, more combinations are tried to account for this relationship.

```
ab_init = AdaBoostClassifier(random_state=seed)
ab = gridsearchcv_fit(ab_init, ab_params, x_train, y_train)
```

Fitting 10 folds for each of 16 candidates, totalling 160 fits

```
[Parallel(n_jobs=-1)]: Done 34 tasks      | elapsed: 2.0min
[Parallel(n_jobs=-1)]: Done 160 out of 160 | elapsed: 8.7min finished
```

```
Search found the best params: {'learning_rate': 0.01, 'n_estimators': 1200}
ROC AUC Train Score: 0.9967975931969577
```

```
ab
```

```
AdaBoostClassifier(algorithm='SAMME.R', base_estimator=None,
    learning_rate=0.01, n_estimators=1200, random_state=0)
```

The AdaBoost model implemented by SciKit-Learn uses the SAMME.R algorithm by default. This algorithm extends the original AdaBoost algorithm by generalizing from the binary classification problem to multi-class problems. The weak learner used by default is the decision tree that splits using the Gini index. The model found the optimal value combination of 1200 trees at a rate of 0.01.

Feature Importances

```
f = plt.figure(figsize=(15,5))
bar_x_axis = range(len(ab.feature_importances_))
plt.bar(bar_x_axis, ab.feature_importances_, width=0.9, color="g",
    tick_label=bar_x_axis)
plt.tick_params(labelsize=10)
plt.show()
```



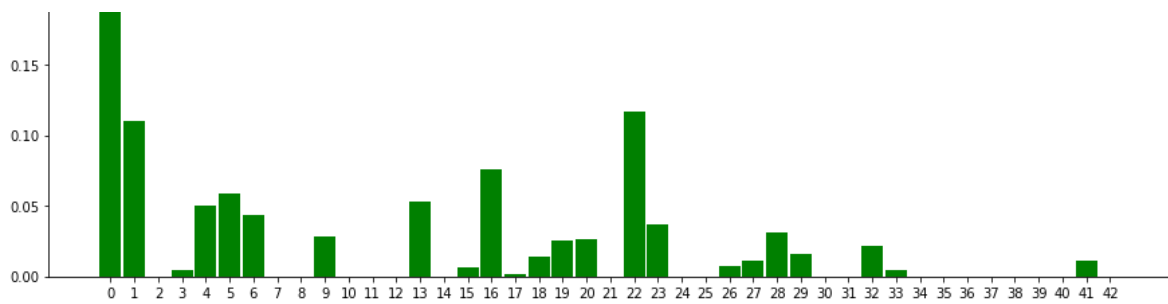


Figure 7: Bar chart of AdaBoost's feature importances. Interestingly, the model has considered a few principle components (e.g. 16, 22) to be relatively more important than those features that correspond to principle components that explain more variance.

Prediction Performance

```
ab_y_pred = ab.predict(x_test)
ab_cm = confusion_matrix(y_true=y_test, y_pred=ab_y_pred)
ab_cm
```

```
array([[58078,  484],
       [ 379, 7059]])
```

The AdaBoost model has predicted 58078 true negatives and 7059 true positives.

```
pr_mnist['AdaBoost'] = precision_recall(ab_cm)
pr_mnist
```

```
{'AdaBoost': (93.58345485880949, 94.90454423232052),
 'Random Forest': (98.6388770735857, 93.53320785157301)}
```

The AdaBoost model has a positive prediction accuracy of 97.06% (precision).

The rate of positive instances that are correctly detected is 89.19% (recall/sensitivity). Unlike the Random Forest, the model is found to be more sensitive than precise.

```
ab_y_score = ab.predict_proba(x_test)[: , 1]
plot_roc(y_test, ab_y_score)
auc_mnist['AdaBoost'] = roc_auc_score(y_true=y_test, y_score=ab_y_score)
print("ROC AUC Test Score: {}".format(auc_mnist['AdaBoost']))
plt.show()
```

```
ROC AUC Test Score: 0.9964689831372104
```

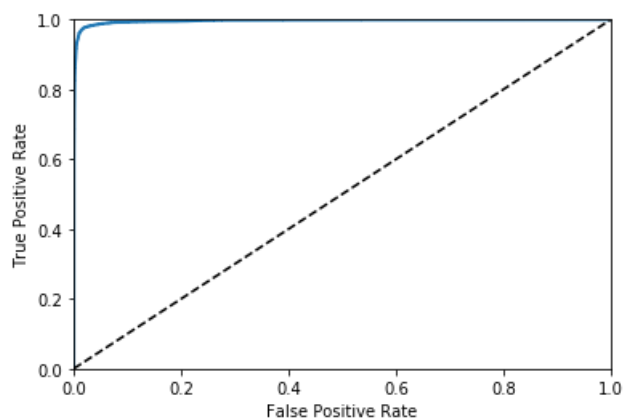


Figure 8: AdaBoost's ROC Curve on the test data.


```
auc_mnist
```

```
{'AdaBoost': 0.9964689831372104, 'Random Forest': 0.9981555217541017}
```

So far, the Random Forest is winning.

XGBoost

```
xg_params = [
    {'n_estimators': [ 800, 1000, 1200, 1400 ], 'learning_rate': [ 0.0001, 0.001, 0.01,
    0.1 ],
    'max_depth': [ 8, 12, 16 ], 'colsample_bytree': [ np.sqrt(p_reduced) / p_reduced, 1
    / 3, 1 ]}
]
```

- `n_estimators`: The number of trees.
- `learning_rate`: Same as AdaBoost.
- `max_depth`: Controls the limit of how many nodes each tree can grow which increases the complexity and flexibility of the model. Theoretically, subtler features should be easier detected with the risk of interpreting noise as information.
- `colsample_bytree`: This hyperparameter is analogous to Random Forest's `max_features` and is chosen for adjustment to provide a member to the stacking methods that can be considered a mix between AdaBoost and Random Forest for balanced diversity. Therefore, similar values are chosen for searching (\sqrt{p}) and $p / 3$ in addition to the 1 value which corresponds to the subset size being equal to the size of the original set of features. This parameter requires a float and so the values must be normalized.

```
xg_init = XGBClassifier(random_state=seed)
xg = gridsearchcv_fit(xg_init, xg_params, x_train, y_train)
```

Fitting 10 folds for each of 144 candidates, totalling 1440 fits

```
[Parallel(n_jobs=-1)]: Done 34 tasks      | elapsed: 25.5s
[Parallel(n_jobs=-1)]: Done 184 tasks     | elapsed: 2.4min
[Parallel(n_jobs=-1)]: Done 434 tasks     | elapsed: 4.9min
[Parallel(n_jobs=-1)]: Done 784 tasks     | elapsed: 11.8min
[Parallel(n_jobs=-1)]: Done 1234 tasks    | elapsed: 30.6min
[Parallel(n_jobs=-1)]: Done 1440 out of 1440 | elapsed: 35.5min finished
```

```
Search found the best params: {'max_depth': 12, 'n_estimators': 1200, 'learning_rate': 0.01, 'colsample_bytree':
0.3333333333333333}
ROC AUC Train Score: 0.998674259961604
```

```
xg
```

```
XGBClassifier(base_score=0.5, booster='gbtree', colsample_bylevel=1,
colsample_bytree=0.3333333333333333, gamma=0, learning_rate=0.01,
max_delta_step=0, max_depth=12, min_child_weight=1, missing=None,
n_estimators=1200, n_jobs=1, nthread=None,
objective='binary:logistic', random_state=0, reg_alpha=0,
reg_lambda=1, scale_pos_weight=1, seed=None, silent=True,
subsample=1)
```

The XGBoost model was fit using the maker's version of a decision tree. A logistic regression model is used as the cost function that is minimized for error convergence. The search found that the optimal model was tuned with feature subset size $m = p / 3$, tree depth of 12, learning rate $\eta = 0.01$, with a number of 1200 trees.

Feature Importances

```
f = plt.figure(figsize=(15,5))
bar_x_axis = range(len(xg.feature_importances_))
plt.bar(bar_x_axis, xg.feature_importances_, width=0.9, color="g")
```

```
plt.bar(bar_x_axis, xg.feature_importances_, width=0.7, color='g',
        tick_label=bar_x_axis)
plt.tick_params(labelsize=10)
plt.show()
```

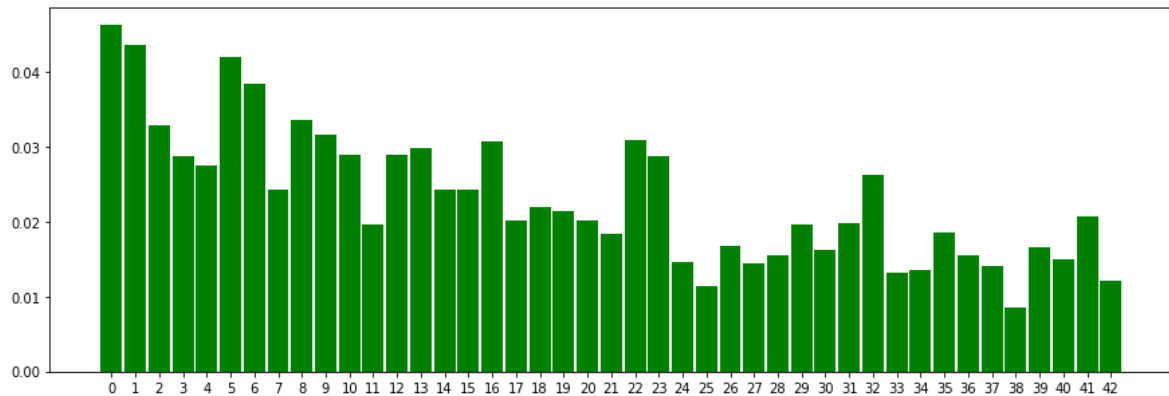


Figure 9: Bar chart of XGBoost's feature importances. Compared to the previous models, XGBoost attributes more equal importance among all features.

Prediction Performance

```
xg_y_pred = xg.predict(x_test)
xg_cm = confusion_matrix(y_true=y_test, y_pred=xg_y_pred)
xg_cm
```

```
/home/cab/.local/lib/python3.5/site-packages/sklearn/preprocessing/label.py:151: DeprecationWarning: The truth value of an empty array is ambiguous. Returning False, but in future this will result in an error. Use `array.size > 0` to check that an array is not empty.
  if diff:
```

```
array([[58440,  122],
       [ 352,  7086]])
```

The XGBoost model has predicted 58440 true negatives and 7086 true positives.

```
pr_mnist['XGBoost'] = precision_recall(xg_cm)
pr_mnist
```

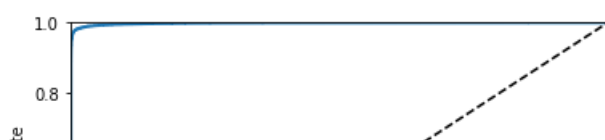
```
{ 'AdaBoost': (93.58345485880949, 94.90454423232052),
  'Random Forest': (98.6388770735857, 93.53320785157301),
  'XGBoost': (98.30743618201998, 95.26754503898897) }
```

The XGBoost model has a positive prediction accuracy of 98.31% (precision).

The rate of positive instances that are correctly detected is 95.27% (recall/sensitivity). Like Random Forest, it is more precise than sensitive, but the difference is smaller.

```
xg_y_score = xg.predict_proba(x_test)[:, 1]
plot_roc(y_test, xg_y_score)
auc_mnist['XGBoost'] = roc_auc_score(y_true=y_test, y_score=xg_y_score)
print("ROC AUC Test Score: {}".format(auc_mnist['XGBoost']))
plt.show()
```

```
ROC AUC Test Score: 0.9982809452784596
```



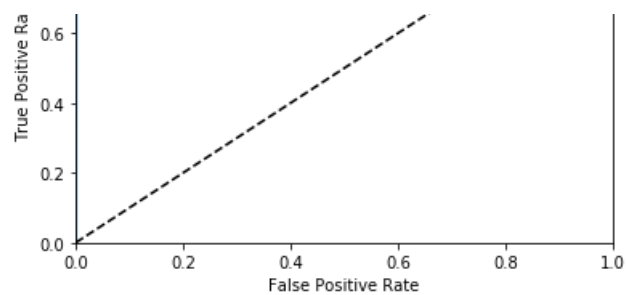


Figure 10: XBoost's ROC Curve on the test data.

```
auc_mnist
```

```
{'AdaBoost': 0.9964689831372104,
 'Random Forest': 0.9981555217541017,
 'XGBoost': 0.9982809452784596}
```

XGBoost has taken a slight lead overall.

Model Stacking

```
rf_oobtrain, rf_oobtest = get_oob(rf, x_train, y_train, x_test, seed)
ab_oobtrain, ab_oobtest = get_oob(ab, x_train, y_train, x_test, seed)
xg_oobtrain, xg_oobtest = get_oob(xg, x_train, y_train, x_test, seed)
print("OoB predictions complete!")
```

```
[Parallel(n_jobs=-1)]: Done 34 tasks      | elapsed: 0.1s
[Parallel(n_jobs=-1)]: Done 184 tasks     | elapsed: 0.3s
[Parallel(n_jobs=-1)]: Done 434 tasks     | elapsed: 0.6s
[Parallel(n_jobs=-1)]: Done 784 tasks     | elapsed: 1.1s
[Parallel(n_jobs=-1)]: Done 1234 tasks    | elapsed: 1.7s
[Parallel(n_jobs=-1)]: Done 1784 tasks    | elapsed: 2.5s
[Parallel(n_jobs=-1)]: Done 2000 out of 2000 | elapsed: 2.9s finished
[Parallel(n_jobs=8)]: Done 34 tasks      | elapsed: 0.0s
[Parallel(n_jobs=8)]: Done 184 tasks     | elapsed: 0.0s
[Parallel(n_jobs=8)]: Done 434 tasks     | elapsed: 0.1s
[Parallel(n_jobs=8)]: Done 784 tasks     | elapsed: 0.1s
[Parallel(n_jobs=8)]: Done 1234 tasks    | elapsed: 0.2s
[Parallel(n_jobs=8)]: Done 1784 tasks    | elapsed: 0.3s
[Parallel(n_jobs=8)]: Done 2000 out of 2000 | elapsed: 0.3s finished
[Parallel(n_jobs=8)]: Done 34 tasks      | elapsed: 0.0s
[Parallel(n_jobs=8)]: Done 184 tasks     | elapsed: 0.2s
[Parallel(n_jobs=8)]: Done 434 tasks     | elapsed: 0.4s
[Parallel(n_jobs=8)]: Done 784 tasks     | elapsed: 0.7s
[Parallel(n_jobs=8)]: Done 1234 tasks    | elapsed: 1.2s
[Parallel(n_jobs=8)]: Done 1784 tasks    | elapsed: 1.7s
[Parallel(n_jobs=8)]: Done 2000 out of 2000 | elapsed: 1.9s finished
[Parallel(n_jobs=-1)]: Done 34 tasks     | elapsed: 0.1s
[Parallel(n_jobs=-1)]: Done 184 tasks    | elapsed: 0.3s
[Parallel(n_jobs=-1)]: Done 434 tasks    | elapsed: 0.6s
[Parallel(n_jobs=-1)]: Done 784 tasks    | elapsed: 1.1s
[Parallel(n_jobs=-1)]: Done 1234 tasks   | elapsed: 1.7s
[Parallel(n_jobs=-1)]: Done 1784 tasks   | elapsed: 2.6s
[Parallel(n_jobs=-1)]: Done 2000 out of 2000 | elapsed: 2.9s finished
[Parallel(n_jobs=8)]: Done 34 tasks      | elapsed: 0.0s
[Parallel(n_jobs=8)]: Done 184 tasks     | elapsed: 0.0s
[Parallel(n_jobs=8)]: Done 434 tasks     | elapsed: 0.1s
[Parallel(n_jobs=8)]: Done 784 tasks     | elapsed: 0.1s
[Parallel(n_jobs=8)]: Done 1234 tasks    | elapsed: 0.2s
[Parallel(n_jobs=8)]: Done 1784 tasks    | elapsed: 0.3s
[Parallel(n_jobs=8)]: Done 2000 out of 2000 | elapsed: 0.3s finished
[Parallel(n_jobs=8)]: Done 34 tasks      | elapsed: 0.0s
[Parallel(n_jobs=8)]: Done 184 tasks     | elapsed: 0.2s
[Parallel(n_jobs=8)]: Done 434 tasks     | elapsed: 0.4s
[Parallel(n_jobs=8)]: Done 784 tasks     | elapsed: 0.7s
[Parallel(n_jobs=8)]: Done 1234 tasks    | elapsed: 1.1s
[Parallel(n_jobs=8)]: Done 1784 tasks    | elapsed: 1.6s
[Parallel(n_jobs=8)]: Done 2000 out of 2000 | elapsed: 1.8s finished
[Parallel(n_jobs=-1)]: Done 34 tasks     | elapsed: 0.1s
```

[illegible]

[illegible]


```
/home/cab/.local/lib/python3.5/site-packages/sklearn/preprocessing/label.py:151: DeprecationWarning: The truth value of an empty array is ambiguous. Returning False, but in future this will result in an error. Use `array.size > 0` to check that an array is not empty.
  if diff:
```

```
x_train_meta = np.concatenate((rf_oobtrain, ab_oobtrain, xg_oobtrain), axis=1)
x_test_meta = np.concatenate((rf_oobtest, ab_oobtest, xg_oobtest), axis=1)
print("Meta learner's input data sizes:\nTrain: {}, Test: {}".format(x_train_meta.shape, x_test_meta.shape))
```

```
Meta learner's input data sizes:
Train: (4000, 3), Test: (66000, 3)
```

- Predictions are performed on cross validated test folds to avoid using the test set for prediction.
- This is achieved by first dividing the training set into 10 distinct folds.
- The model is trained on 9 folds, and predictions are made on the remaining fold as well as on the test set.
- This is repeated 9 more times, each time using a different fold to predict on where the other 9 folds are used for training. Each time predictions are also made on the test set.
- The predictions of each fold are gathered to form a single dataset known as the out-of-bag training set (_oobtrain).
- The predictions made on the test set (by 10 versions of the model trained on 9 folds) are merged by taking the mean of all predictions made on each instance resulting in a single dataset known as the out-of-bag test set (_oobtest).
- The out-of-bag training set is fed as the input data for a meta learner stacking model.
- The out-of-bag test set is used for producing the test scores.

Simple Majority Vote

```
mv_y_scores = x_test_meta.mean(axis=1).reshape(-1,1)
mv_y_pred = np.around(mv_y_scores)
```

Tallies a simple vote for each instance, a majority of 1's predicts a 1, else a 0 is predicted. The scores represent the probability of a 1, e.g. if there are 2 out of 3 votes for a 1 to be predicted, the score is 0.67.

Prediction Performance

```
mv_cm = confusion_matrix(y_true=y_test, y_pred=mv_y_pred)
mv_cm
```

```
array([[58435, 127],
       [ 400, 7038]])
```

The Majority Vote has predicted 58435 true negatives and 7038 true positives.

```
pr_mnist['Majority Vote'] = precision_recall(mv_cm)
pr_mnist
```

```
{'AdaBoost': (93.58345485880949, 94.90454423232052),
 'Majority Vote': (98.2274947662247, 94.62221027157838),
 'Random Forest': (98.6388770735857, 93.53320785157301),
 'XGBoost': (98.30743618201998, 95.26754503898897)}
```

The Majority Vote model has a positive prediction accuracy of 98.23% (precision). The rate of positive instances that are correctly detected is 94.62% (recall/sensitivity). Being that it's nature is to average the other model's predictions, it is no surprise that it has found its way to the middle of the pack in precision and recall.

```

plot_roc(y_test, mv_y_scores)
auc_mnist['Majority Vote'] = roc_auc_score(y_true=y_test, y_score=mv_y_scores)
print("ROC AUC Test Score: {}".format(auc_mnist['Majority Vote']))
plt.show()

```

ROC AUC Test Score: 0.9797359353447191

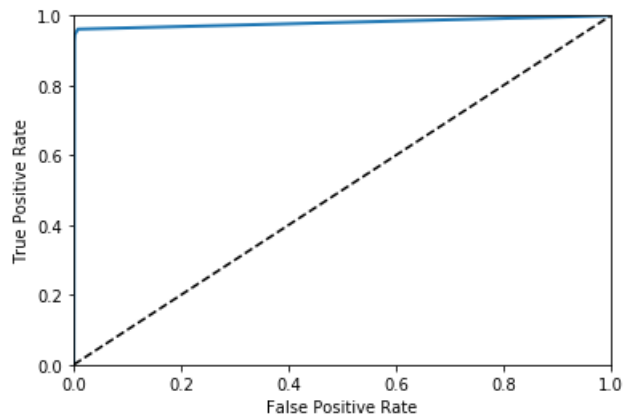


Figure 11: ROC Curve on the test data from the Majority Vote model.

```
auc_mnist
```

```

{'AdaBoost': 0.9964689831372104,
 'Majority Vote': 0.9797359353447191,
 'Random Forest': 0.9981555217541017,
 'XGBoost': 0.9982809452784596}

```

The model appears to be overfitting, it has placed last. Treating each of the other model's equally to each other appears to be hampering performance.

Meta Learner Random Forest

```

metarf_params = [
    {'n_estimators': [ 10, 20, 50, 100, 200, 400 ]}
]

```

max_features has been tuned to a constant of sqrt(p) considering there are only 3 features. A subset size of 1 feature is assumed to be useless.

```

metarf_init = RandomForestClassifier(max_features="sqrt", random_state=seed, n_jobs=n_jobs,
    verbose=1)
metarf = gridsearchcv_fit(metarf_init, metarf_params, x_train_meta, y_train)

```

Fitting 10 folds for each of 6 candidates, totalling 60 fits

```

[Parallel(n_jobs=-1)]: Done 6 out of 10 | elapsed: 0.0s remaining: 0.0s
[Parallel(n_jobs=-1)]: Done 6 out of 10 | elapsed: 0.0s remaining: 0.0s
[Parallel(n_jobs=-1)]: Done 10 out of 10 | elapsed: 0.0s finished
[Parallel(n_jobs=-1)]: Done 6 out of 10 | elapsed: 0.0s remaining: 0.0s
[Parallel(n_jobs=-1)]: Done 10 out of 10 | elapsed: 0.0s finished
[Parallel(n_jobs=-1)]: Done 10 out of 10 | elapsed: 0.0s finished
[Parallel(n_jobs=-1)]: Done 6 out of 10 | elapsed: 0.0s remaining: 0.0s
[Parallel(n_jobs=-1)]: Done 6 out of 10 | elapsed: 0.0s remaining: 0.0s
[Parallel(n_jobs=-1)]: Done 6 out of 10 | elapsed: 0.0s remaining: 0.0s
[Parallel(n_jobs=-1)]: Done 10 out of 10 | elapsed: 0.0s finished
[Parallel(n_jobs=-1)]: Done 10 out of 10 | elapsed: 0.0s finished
[Parallel(n_jobs=-1)]: Done 10 out of 10 | elapsed: 0.0s finished
[Parallel(n_jobs=-1)]: Done 6 out of 10 | elapsed: 0.0s remaining: 0.0s
[Parallel(n_jobs=-1)]: Done 10 out of 10 | elapsed: 0.0s finished
[Parallel(n_jobs=-1)]: Done 6 out of 10 | elapsed: 0.0s remaining: 0.0s
[Parallel(n_jobs=-1)]: Done 10 out of 10 | elapsed: 0.0s finished

```


[illegible]

[illegible]

[illegible]

[illegible]

```
[Parallel(n_jobs=8)]: Done 400 out of 400 | elapsed: 0.1s finished
[Parallel(n_jobs=8)]: Done 400 out of 400 | elapsed: 0.1s finished
[Parallel(n_jobs=8)]: Done 400 out of 400 | elapsed: 0.1s finished
[Parallel(n_jobs=8)]: Done 34 tasks | elapsed: 0.0s
[Parallel(n_jobs=8)]: Done 34 tasks | elapsed: 0.0s
[Parallel(n_jobs=8)]: Done 34 tasks | elapsed: 0.0s
[Parallel(n_jobs=-1)]: Done 34 tasks | elapsed: 0.0s
[Parallel(n_jobs=-1)]: Done 34 tasks | elapsed: 0.0s
[Parallel(n_jobs=8)]: Done 184 tasks | elapsed: 0.1s
[Parallel(n_jobs=8)]: Done 184 tasks | elapsed: 0.1s
[Parallel(n_jobs=8)]: Done 184 tasks | elapsed: 0.1s
[Parallel(n_jobs=8)]: Done 400 out of 400 | elapsed: 0.2s finished
[Parallel(n_jobs=8)]: Done 400 out of 400 | elapsed: 0.2s finished
[Parallel(n_jobs=-1)]: Done 184 tasks | elapsed: 0.2s
[Parallel(n_jobs=8)]: Done 400 out of 400 | elapsed: 0.2s finished
[Parallel(n_jobs=-1)]: Done 184 tasks | elapsed: 0.2s
[Parallel(n_jobs=-1)]: Done 400 out of 400 | elapsed: 0.3s finished
[Parallel(n_jobs=-1)]: Done 400 out of 400 | elapsed: 0.3s finished
[Parallel(n_jobs=-1)]: Done 34 tasks | elapsed: 0.0s
[Parallel(n_jobs=8)]: Done 34 tasks | elapsed: 0.0s
[Parallel(n_jobs=8)]: Done 34 tasks | elapsed: 0.0s
[Parallel(n_jobs=-1)]: Done 34 tasks | elapsed: 0.0s
[Parallel(n_jobs=8)]: Done 184 tasks | elapsed: 0.0s
[Parallel(n_jobs=8)]: Done 184 tasks | elapsed: 0.0s
[Parallel(n_jobs=8)]: Done 400 out of 400 | elapsed: 0.1s finished
[Parallel(n_jobs=8)]: Done 400 out of 400 | elapsed: 0.1s finished
[Parallel(n_jobs=-1)]: Done 184 tasks | elapsed: 0.1s
[Parallel(n_jobs=8)]: Done 34 tasks | elapsed: 0.0s
[Parallel(n_jobs=8)]: Done 34 tasks | elapsed: 0.0s
[Parallel(n_jobs=-1)]: Done 184 tasks | elapsed: 0.1s
[Parallel(n_jobs=8)]: Done 184 tasks | elapsed: 0.1s
[Parallel(n_jobs=8)]: Done 184 tasks | elapsed: 0.1s
[Parallel(n_jobs=8)]: Done 400 out of 400 | elapsed: 0.1s finished
[Parallel(n_jobs=8)]: Done 400 out of 400 | elapsed: 0.1s finished
[Parallel(n_jobs=-1)]: Done 400 out of 400 | elapsed: 0.3s finished
[Parallel(n_jobs=-1)]: Done 400 out of 400 | elapsed: 0.3s finished
[Parallel(n_jobs=8)]: Done 34 tasks | elapsed: 0.0s
[Parallel(n_jobs=8)]: Done 34 tasks | elapsed: 0.0s
[Parallel(n_jobs=8)]: Done 184 tasks | elapsed: 0.0s
[Parallel(n_jobs=8)]: Done 184 tasks | elapsed: 0.0s
[Parallel(n_jobs=8)]: Done 400 out of 400 | elapsed: 0.1s finished
[Parallel(n_jobs=8)]: Done 400 out of 400 | elapsed: 0.1s finished
[Parallel(n_jobs=8)]: Done 34 tasks | elapsed: 0.0s
[Parallel(n_jobs=8)]: Done 34 tasks | elapsed: 0.0s
[Parallel(n_jobs=8)]: Done 184 tasks | elapsed: 0.0s
[Parallel(n_jobs=8)]: Done 184 tasks | elapsed: 0.0s
[Parallel(n_jobs=8)]: Done 400 out of 400 | elapsed: 0.1s finished
[Parallel(n_jobs=8)]: Done 400 out of 400 | elapsed: 0.1s finished
[Parallel(n_jobs=8)]: Done 400 out of 400 | elapsed: 0.1s finished
[Parallel(n_jobs=-1)]: Done 60 out of 60 | elapsed: 4.8s finished
[Parallel(n_jobs=-1)]: Done 6 out of 10 | elapsed: 0.0s remaining: 0.0s
[Parallel(n_jobs=-1)]: Done 10 out of 10 | elapsed: 0.0s finished
```

Search found the best params: {'n_estimators': 10}
ROC AUC Train Score: 0.9781167061338397

metarf

```
RandomForestClassifier(bootstrap=True, class_weight=None, criterion='gini',
                        max_depth=None, max_features='sqrt', max_leaf_nodes=None,
                        min_impurity_decrease=0.0, min_impurity_split=None,
                        min_samples_leaf=1, min_samples_split=2,
                        min_weight_fraction_leaf=0.0, n_estimators=10, n_jobs=-1,
                        oob_score=False, random_state=0, verbose=1, warm_start=False)
```

The search found that overfitting happens fast as the minimum value of 10 trees is chosen to be optimal.

Prediction Performance

```
metarf_y_pred = metarf.predict(x_test_meta)
metarf_cm = confusion_matrix(y_true=y_test, y_pred=metarf_y_pred)
metarf_cm
```

```
[Parallel(n_jobs=8)]: Done 6 out of 10 | elapsed: 0.0s remaining: 0.0s
```

```
[Parallel(n_jobs=8)]: Done 10 out of 10 | elapsed: 0.0s finished
```

```
array([[58435, 127],  
       [ 350, 7088]])
```

The Meta Forest model has predicted 58435 true negatives and 7088 true positives.

```
pr_mnist['Meta-Forest'] = precision_recall(metarf_cm)  
pr_mnist
```

```
{'AdaBoost': (93.58345485880949, 94.90454423232052),  
 'Majority Vote': (98.2274947662247, 94.62221027157838),  
 'Meta-Forest': (98.23977823977825, 95.2944339876311),  
 'Random Forest': (98.6388770735857, 93.53320785157301),  
 'XGBoost': (98.30743618201998, 95.26754503898897)}
```

The Meta Forest model has a positive prediction accuracy of 98.24% (precision).

The rate of positive instances that are correctly detected is 95.29% (recall/sensitivity). The Meta Forest slightly outperforms the Majority Vote in precision, while also outperforming the Random Forest in recall.

```
metarf_y_score = metarf.predict_proba(x_test_meta)[: , 1]  
plot_roc(y_test, metarf_y_score)  
auc_mnist['Meta-Forest'] = roc_auc_score(y_true=y_test, y_score=metarf_y_score)  
print("ROC AUC Test Score: {}".format(auc_mnist['Meta-Forest']))  
plt.show()
```

```
[Parallel(n_jobs=8)]: Done 6 out of 10 | elapsed: 0.0s remaining: 0.0s  
[Parallel(n_jobs=8)]: Done 10 out of 10 | elapsed: 0.0s finished
```

```
ROC AUC Test Score: 0.9794651908780632
```

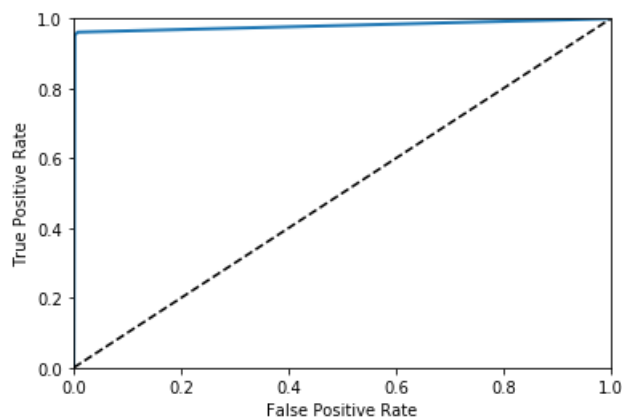


Figure 12: ROC Curve on the test data from the Meta Forest.

```
auc_mnist
```

```
{'AdaBoost': 0.9964689831372104,  
 'Majority Vote': 0.9797359353447191,  
 'Meta-Forest': 0.9794651908780632,  
 'Random Forest': 0.9981555217541017,  
 'XGBoost': 0.9982809452784596}
```

Unfortunately, decorrelating the influence of each of the lower layer models doesn't appear to improve overall performance over the Majority Vote model.

Meta Learner XGBoost

```
metaxg_params = [
    {'n_estimators': [ 1200, 1400, 1600, 1800 ], 'learning_rate': [ 1e-12, 1e-10, 1e-08
, 1e-06 ],
    'max_depth': [ 1, 2 ], 'colsample_bytree': [ 1 / 3, 2 / 3, 1 ] }
]
```

With only 3 features, it is expected that the optimal model might require more trees (and thus slower learning rates) than the non-meta learner model.

colsample_bytree is considered useful in this model because this algorithm involves weights. That is, even if only one feature is used to split, the overall weight attributed to the resulting tree will provide some assurance of its usefulness.

```
metaxg_init = XGBClassifier(random_state=seed)
metaxg = gridsearchcv_fit(metaxg_init, metaxg_params, x_train_meta, y_train)
```

Fitting 10 folds for each of 96 candidates, totalling 960 fits

```
[Parallel(n_jobs=-1)]: Done 34 tasks      | elapsed:    2.4s
[Parallel(n_jobs=-1)]: Done 184 tasks     | elapsed:   13.2s
[Parallel(n_jobs=-1)]: Done 434 tasks     | elapsed:   32.6s
[Parallel(n_jobs=-1)]: Done 784 tasks     | elapsed:   1.0min
[Parallel(n_jobs=-1)]: Done 960 out of 960 | elapsed:   1.3min finished
```

```
Search found the best params: {'max_depth': 1, 'n_estimators': 1800, 'learning_rate': 1e-10, 'colsample_bytree':
0.3333333333333333}
ROC AUC Train Score: 0.9805043465832778
```

metaxg

```
XGBClassifier(base_score=0.5, booster='gbtree', colsample_bylevel=1,
    colsample_bytree=0.3333333333333333, gamma=0, learning_rate=1e-10,
    max_delta_step=0, max_depth=1, min_child_weight=1, missing=None,
    n_estimators=1800, n_jobs=1, nthread=None,
    objective='binary:logistic', random_state=0, reg_alpha=0,
    reg_lambda=1, scale_pos_weight=1, seed=None, silent=True,
    subsample=1)
```

The search found that restricting flexibility to the additive model (max_depth = 1) prevented overfitting. Unexpectedly, the model that subsetted the feature space to 0.3 = 1 feature was found to be the best. The maximum number of 1800 trees at a learning rate of 1e-10 was found which gives reason to re-run the search with higher values. However, it was noticed that increasing further did not decrease performance.

Prediction Performance

```
metaxg_y_pred = metaxg.predict(x_test_meta)
metaxg_cm = confusion_matrix(y_true=y_test, y_pred=metaxg_y_pred)
metaxg_cm
```

```
/home/cab/.local/lib/python3.5/site-packages/sklearn/preprocessing/label.py:151: DeprecationWarning: The truth va
lue of an empty array is ambiguous. Returning False, but in future this will result in an error. Use `array.size
> 0` to check that an array is not empty.
    if diff:
```

```
array([[58435,   127],
       [   400,  7038]])
```

The Meta XGBoost model has predicted 58435 true negatives and 7038 true positives. This is identical to the Majority Vote matrix.

```
pr_mnist['Meta-XGBoost'] = precision_recall(metaxg_cm)
pr_mnist
```

```
{'AdaBoost': (93.58345485880949, 94.90454423232052),
'Majority Vote': (98.2274947662247, 94.62221027157838),
'Meta-Forest': (98.23977823977825, 95.2944339876311),
'Meta-XGBoost': (98.2274947662247, 94.62221027157838),
'Random Forest': (98.6388770735857, 93.53320785157301),
'XGBoost': (98.30743618201998, 95.26754503898897)}
```

The Meta XGBoost model has a positive prediction accuracy of 98.23% (precision). The rate of positive instances that are correctly detected is 94.62% (recall/sensitivity). As mentioned, this is identical to the Majority Vote results.

```
metaxg_y_score = metaxg.predict_proba(x_test_meta)[: , 1]
plot_roc(y_test, metaxg_y_score)
auc_mnist['Meta-XGBoost'] = roc_auc_score(y_true=y_test, y_score=metaxg_y_score)
print("ROC AUC Test Score: {}".format(auc_mnist['Meta-XGBoost']))
plt.show()
```

ROC AUC Test Score: 0.9797359353447191

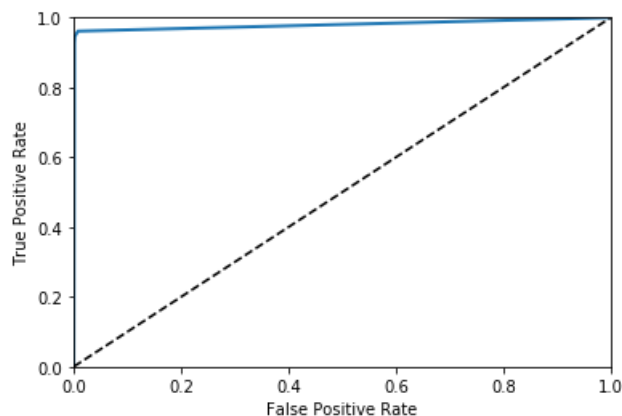


Figure 13: ROC Curve on the test data from the Meta XGBoost.

```
auc_mnist
```

```
{'AdaBoost': 0.9964689831372104,
'Majority Vote': 0.9797359353447191,
'Meta-Forest': 0.9794651908780632,
'Meta-XGBoost': 0.9797359353447191,
'Random Forest': 0.9981555217541017,
'XGBoost': 0.9982809452784596}
```

Perhaps this model has learned that the best stacking model on this dataset is the Majority Vote. Overall, the stacking models appear to be beat by the non-stacking models. This is a similar result to the performance of non-linear models compared to linear model when the data is linear. Essentially, these models are trained to identify an image of a straight line (a handwritten number 1), so it could be reasonably argued that models with less flexibility are expected to perform better. The next dataset should provide additional perspective, as its underlying function is expected to be complex and non-linear.

Sheep Dataset Preparation

```
sheep = loadmat('sheep_10s_ws/sheep_10s_ws.mat')
sheep.keys()
```

```
dict_keys(['Ps2feature10s', 'sheep16th10sfeature', '__version__', '__header__', 'Ys2feature10s', 'Gs2feature10s',
'__globals__'])
```


The Sheep data set is collection of 4 data sets, 1 for each sheep and then one that contains all others. The latter is the one used for this analysis. Each instance represents a 10 second window of measurements collected from accelerometers attached to sheep that senses its movement. The movement is then classified as either walking, standing or grazing. These classes are converted for the binary classification models where 1 represents grazing, and 0 represents not grazing (walking or standing).

```
dataset = sheep['sheep16th10sfeature']
dataset.shape
```

```
(5176, 56)
```

The learning methods will be tested for performance in binary classification. Therefore, the original 10 class label is converted into a 2 class label: 1 or not-1. Also, the high dimensions motivate reduction methods such as PCA.

```
data = dataset[:, :-1]
label = (dataset[:, -1] == 3).astype(int)
n, p = data.shape
classes, counts = np.unique(label, return_counts=True)
print("n = {}\np = {}\nclass, count:\n{}".format(n, p, list(zip(classes, counts))))
```

```
n = 5176
p = 55
class, count:
[(0, 1815), (1, 3361)]
```

```
plt.pie(counts, labels=classes)
plt.show()
```

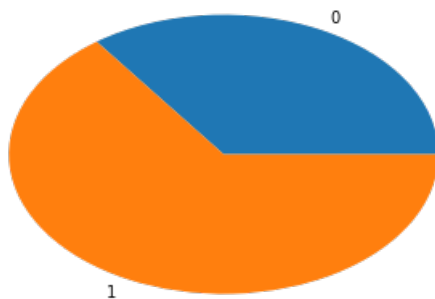


Figure 14: Proportion of classes in this dataset is relatively balanced compared to the MNIST dataset.

Split data into train/test

```
x_train, x_test, y_train, y_test = train_test_split(data, label, train_size=4000, random_state=seed)
y_train, y_test = y_train.ravel(), y_test.ravel()
print("n_train, n_test: {0}, {1}".format(len(y_train), len(y_test)))
```

```
n_train, n_test: 4000, 1176
```

The training size used is equal to that of the previous dataset and so the test set is far smaller.

Correlation Heatmap

```
f, ax = plt.subplots(figsize=(18, 18))
```

```
corr = pd.DataFrame(x_train).corr()
hm = sns.heatmap(round(corr,2), annot=False, ax=ax, cmap="coolwarm",fmt='.2f')
f.subplots_adjust(top=0.93)
```

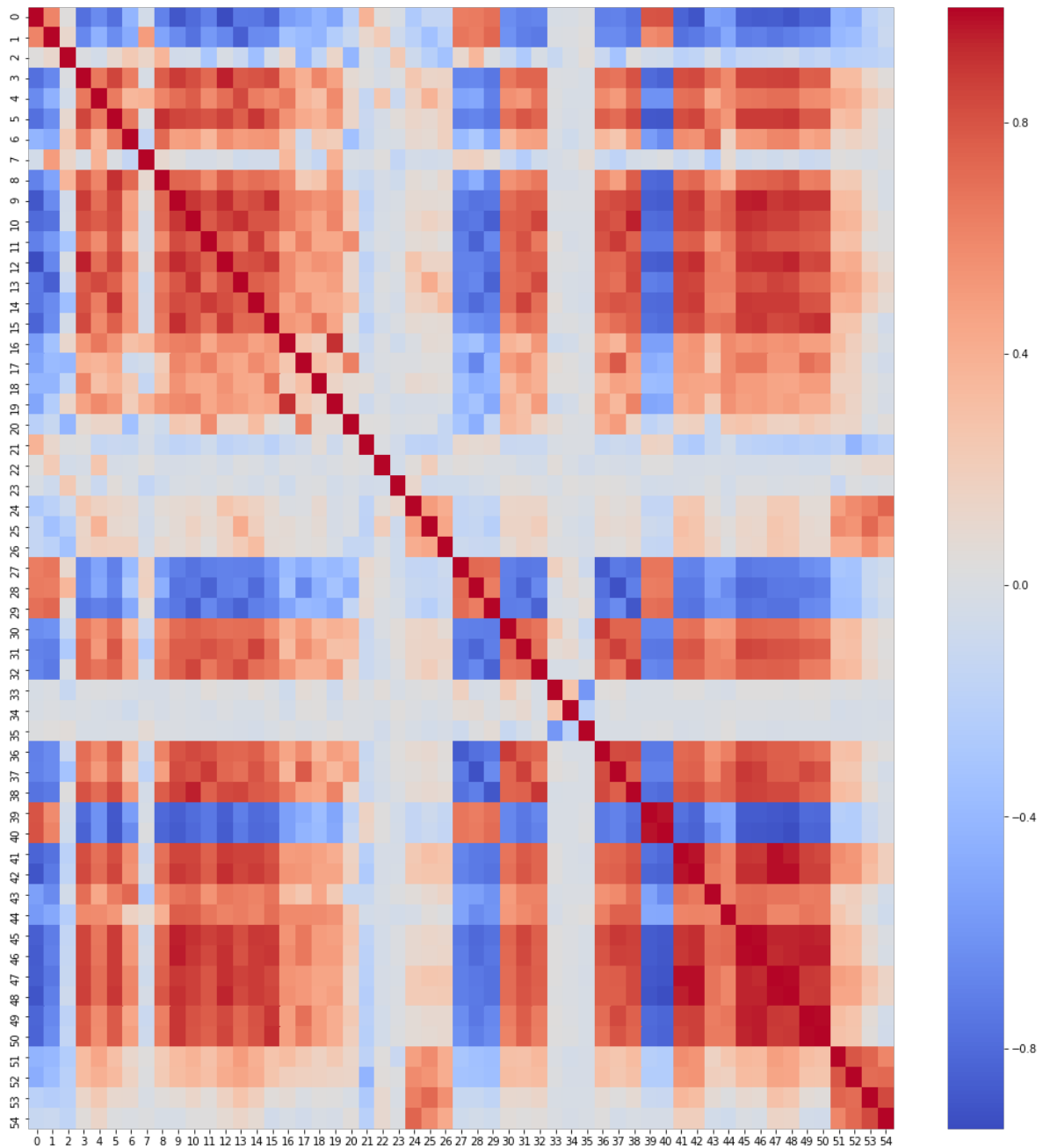


Figure 15: Like the MNIST heatmap, this correlation heatmap of the features also reveals a pattern. In this case, multiple groups of adjacent features can be seen as highly correlated manifesting as approximately 8-10 square-shaped groups of red or blue squares. This reflects the ordering of the features during data collection as accelerometers consist of multiple dimensions to capture spatial data. Further, the heavier degrees of correlation are much more frequent in this case. Apart from a few exceptions (features 20-26, 33-35 and 51-54), almost all features show high correlation with almost all other features. This suggests that the Random Forest should outperform the boosting models.

Principle Component Analysis

```
pca_init = PCA(n_components=0.95)
x_reduced1 = pca_init.fit_transform(x_train)
print("Number of Dimensions:\n{0:3d} - original\n{1:3d} - reduced ({2} dimensions
eliminated) "
      .format(p, x_reduced1.shape[1], p - x_reduced1.shape[1]))
```

```
Number of Dimensions:
55 - original
```

5 - reduced (50 dimensions eliminated)

```
plt.plot(np.cumsum(pca_init.explained_variance_ratio_))
plt.xlabel("Number of Dimensions")
plt.ylabel("Proportion of Variance Explained")
plt.show()
```

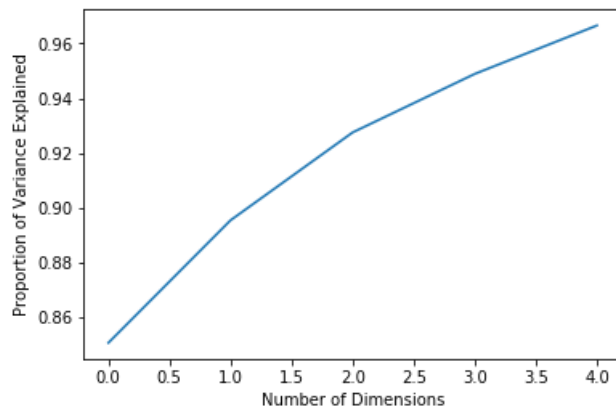


Figure 16: Proportion of variance explained plot shows an elbow at 2 dimensions.

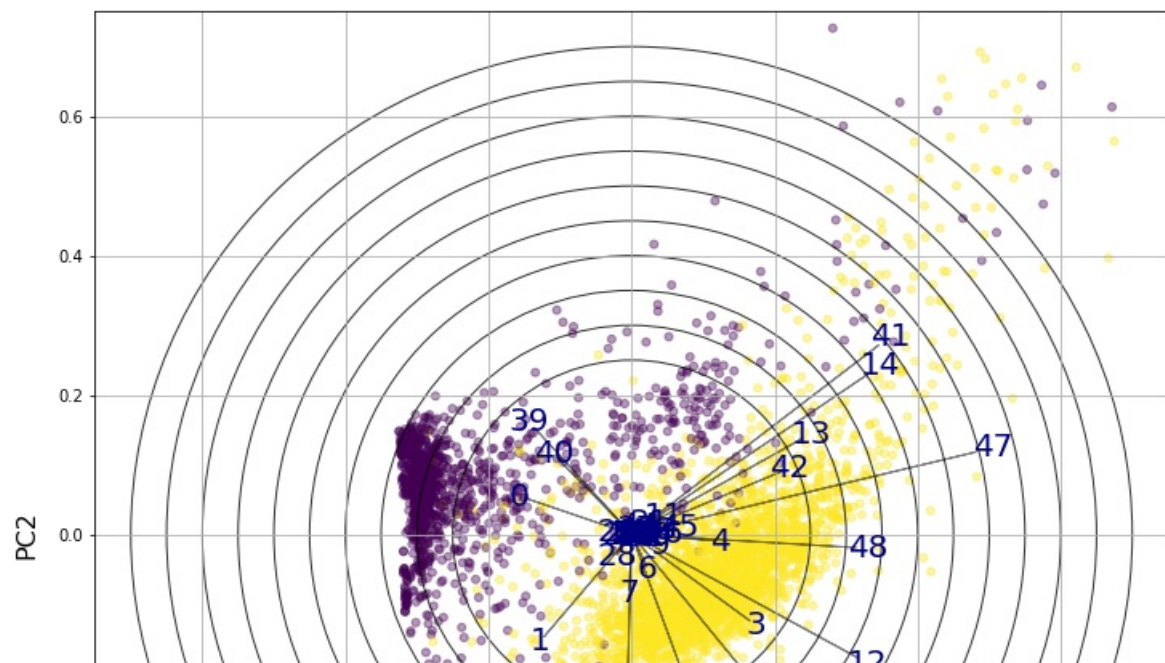
PCA Biplot

```
pca_final = PCA(n_components=2)
x_reduced2 = pca_final.fit_transform(x_train)
pc1evr, pc2evr = np.multiply(pca_final.explained_variance_ratio_[0:2], np.full(2, 100))
print("Proportion of Variance Explained:\n{:.2f}% - PC1\n{:.2f}% - PC2\n{:.2f}% - PC1\nand PC2"
      .format(pc1evr, pc2evr, pc1evr + pc2evr))
```

Proportion of Variance Explained:
85.06% - PC1
4.48% - PC2
89.54% - PC1 and PC2

Another PCA is fit for 2 dimensions before producing the biplot.

```
biplot('sheep', pca_final, x_reduced2, y_train)
```



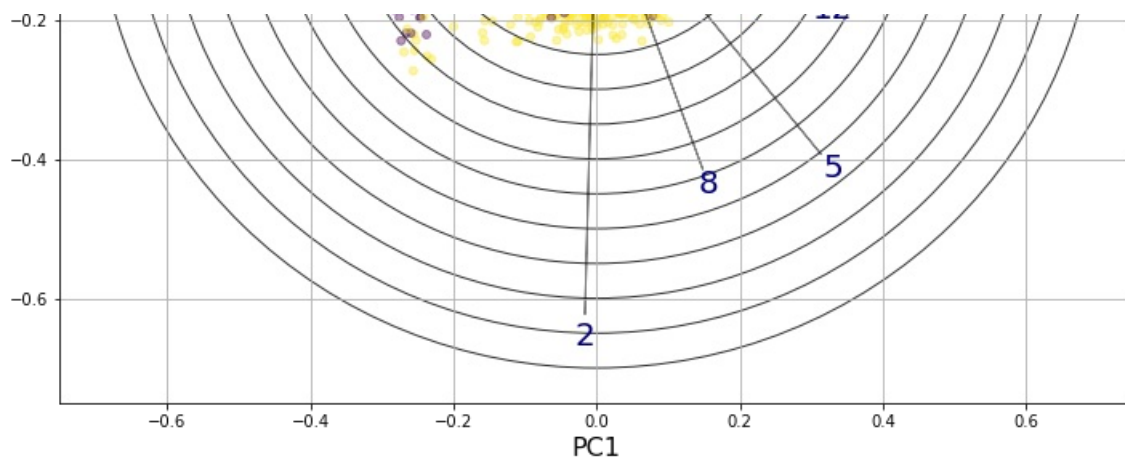


Figure 17: Biplot of all of the two principle components in the final PCA transform. The depicted components account for 89.54% of the variance. Approximately 8-10 vectors appear to have far greater magnitudes than all others, further reflecting the highly correlated nature of the data. The directions of these vectors tend to similar directions, ranging from top right to bottom. The scatter of the data points projected on these dimensions reveals the two classes can be roughly clustered. The alpha density of their colours shows that the clusters centres are well separated. Given these characteristics, the ensembling methods that involve feature decorrelation are expected to outperform others in binary classification.

Model Training

- The number of features $p = 52$ doesn't necessitate the need to reduce dimensions for acceptable training times.

```
auc_sheep = {}
pr_sheep = {}
```

Random Forest

```
rf_params = [
    {'n_estimators': [ 1500, 1750, 2000, 2250 ], 'max_features': [ "sqrt", int(p / 3) ]}
]
```

```
rf_init = RandomForestClassifier(random_state=seed, n_jobs=n_jobs, verbose=1)
rf = gridsearchcv_fit(rf_init, rf_params, x_train, y_train)
```

Fitting 10 folds for each of 8 candidates, totalling 80 fits

```
[Parallel(n_jobs=-1)]: Done 34 tasks      | elapsed: 0.3s
[Parallel(n_jobs=-1)]: Done 34 tasks      | elapsed: 0.3s
[Parallel(n_jobs=-1)]: Done 34 tasks      | elapsed: 0.3s
[Parallel(n_jobs=-1)]: Done 34 tasks      | elapsed: 0.4s
[Parallel(n_jobs=-1)]: Done 34 tasks      | elapsed: 0.6s
[Parallel(n_jobs=-1)]: Done 34 tasks      | elapsed: 0.5s
[Parallel(n_jobs=-1)]: Done 34 tasks      | elapsed: 0.6s
[Parallel(n_jobs=-1)]: Done 34 tasks      | elapsed: 0.5s
[Parallel(n_jobs=-1)]: Done 184 tasks     | elapsed: 1.8s
[Parallel(n_jobs=-1)]: Done 184 tasks     | elapsed: 2.3s
[Parallel(n_jobs=-1)]: Done 184 tasks     | elapsed: 2.4s
[Parallel(n_jobs=-1)]: Done 184 tasks     | elapsed: 2.5s
[Parallel(n_jobs=-1)]: Done 184 tasks     | elapsed: 2.9s
[Parallel(n_jobs=-1)]: Done 184 tasks     | elapsed: 2.9s
[Parallel(n_jobs=-1)]: Done 184 tasks     | elapsed: 3.1s
[Parallel(n_jobs=-1)]: Done 184 tasks     | elapsed: 3.0s
[Parallel(n_jobs=-1)]: Done 434 tasks     | elapsed: 5.6s
[Parallel(n_jobs=-1)]: Done 434 tasks     | elapsed: 6.2s
[Parallel(n_jobs=-1)]: Done 434 tasks     | elapsed: 6.4s
[Parallel(n_jobs=-1)]: Done 434 tasks     | elapsed: 6.7s
[Parallel(n_jobs=-1)]: Done 434 tasks     | elapsed: 6.8s
[Parallel(n_jobs=-1)]: Done 434 tasks     | elapsed: 6.6s
[Parallel(n_jobs=-1)]: Done 434 tasks     | elapsed: 6.8s
[Parallel(n_jobs=-1)]: Done 434 tasks     | elapsed: 6.8s
```

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

```
[Parallel(n_jobs=8)]: Done 2250 out of 2250 | elapsed: 1.3s finished
[Parallel(n_jobs=8)]: Done 1234 tasks | elapsed: 0.3s
[Parallel(n_jobs=8)]: Done 2250 out of 2250 | elapsed: 1.3s finished
[Parallel(n_jobs=8)]: Done 2250 out of 2250 | elapsed: 1.4s finished
[Parallel(n_jobs=8)]: Done 2250 out of 2250 | elapsed: 1.4s finished
[Parallel(n_jobs=8)]: Done 2250 out of 2250 | elapsed: 1.2s finished
[Parallel(n_jobs=8)]: Done 2250 out of 2250 | elapsed: 1.2s finished
[Parallel(n_jobs=8)]: Done 34 tasks | elapsed: 0.0s
[Parallel(n_jobs=8)]: Done 2250 out of 2250 | elapsed: 1.2s finished
[Parallel(n_jobs=8)]: Done 34 tasks | elapsed: 0.0s
[Parallel(n_jobs=8)]: Done 34 tasks | elapsed: 0.0s
[Parallel(n_jobs=8)]: Done 34 tasks | elapsed: 0.0s
[Parallel(n_jobs=8)]: Done 184 tasks | elapsed: 0.1s
[Parallel(n_jobs=8)]: Done 184 tasks | elapsed: 0.1s
[Parallel(n_jobs=8)]: Done 34 tasks | elapsed: 0.0s
[Parallel(n_jobs=8)]: Done 34 tasks | elapsed: 0.0s
[Parallel(n_jobs=8)]: Done 1784 tasks | elapsed: 0.5s
[Parallel(n_jobs=8)]: Done 184 tasks | elapsed: 0.1s
[Parallel(n_jobs=8)]: Done 184 tasks | elapsed: 0.1s
[Parallel(n_jobs=8)]: Done 184 tasks | elapsed: 0.1s
[Parallel(n_jobs=8)]: Done 184 tasks | elapsed: 0.1s
[Parallel(n_jobs=8)]: Done 184 tasks | elapsed: 0.1s
[Parallel(n_jobs=8)]: Done 434 tasks | elapsed: 0.2s
[Parallel(n_jobs=8)]: Done 434 tasks | elapsed: 0.2s
[Parallel(n_jobs=8)]: Done 2250 out of 2250 | elapsed: 0.7s finished
[Parallel(n_jobs=8)]: Done 434 tasks | elapsed: 0.3s
[Parallel(n_jobs=8)]: Done 434 tasks | elapsed: 0.3s
[Parallel(n_jobs=8)]: Done 434 tasks | elapsed: 0.3s
[Parallel(n_jobs=8)]: Done 434 tasks | elapsed: 0.3s
[Parallel(n_jobs=8)]: Done 434 tasks | elapsed: 0.3s
[Parallel(n_jobs=8)]: Done 34 tasks | elapsed: 0.0s
[Parallel(n_jobs=8)]: Done 784 tasks | elapsed: 0.4s
[Parallel(n_jobs=8)]: Done 184 tasks | elapsed: 0.1s
[Parallel(n_jobs=8)]: Done 784 tasks | elapsed: 0.5s
[Parallel(n_jobs=8)]: Done 784 tasks | elapsed: 0.5s
[Parallel(n_jobs=8)]: Done 784 tasks | elapsed: 0.5s
[Parallel(n_jobs=8)]: Done 784 tasks | elapsed: 0.5s
[Parallel(n_jobs=8)]: Done 784 tasks | elapsed: 0.5s
[Parallel(n_jobs=8)]: Done 434 tasks | elapsed: 0.3s
[Parallel(n_jobs=8)]: Done 1234 tasks | elapsed: 0.7s
[Parallel(n_jobs=8)]: Done 1234 tasks | elapsed: 0.7s
[Parallel(n_jobs=8)]: Done 1234 tasks | elapsed: 0.8s
[Parallel(n_jobs=8)]: Done 1234 tasks | elapsed: 0.7s
[Parallel(n_jobs=8)]: Done 1234 tasks | elapsed: 0.7s
[Parallel(n_jobs=8)]: Done 784 tasks | elapsed: 0.5s
[Parallel(n_jobs=8)]: Done 1234 tasks | elapsed: 0.8s
[Parallel(n_jobs=8)]: Done 1234 tasks | elapsed: 0.8s
[Parallel(n_jobs=8)]: Done 1784 tasks | elapsed: 1.1s
[Parallel(n_jobs=8)]: Done 1784 tasks | elapsed: 1.1s
[Parallel(n_jobs=8)]: Done 1234 tasks | elapsed: 0.8s
[Parallel(n_jobs=8)]: Done 1784 tasks | elapsed: 1.1s
[Parallel(n_jobs=8)]: Done 1784 tasks | elapsed: 1.1s
[Parallel(n_jobs=8)]: Done 1784 tasks | elapsed: 1.1s
[Parallel(n_jobs=8)]: Done 1784 tasks | elapsed: 1.1s
[Parallel(n_jobs=8)]: Done 1784 tasks | elapsed: 1.1s
[Parallel(n_jobs=8)]: Done 2250 out of 2250 | elapsed: 1.4s finished
[Parallel(n_jobs=8)]: Done 2250 out of 2250 | elapsed: 1.3s finished
[Parallel(n_jobs=8)]: Done 2250 out of 2250 | elapsed: 1.4s finished
[Parallel(n_jobs=8)]: Done 2250 out of 2250 | elapsed: 1.4s finished
[Parallel(n_jobs=8)]: Done 2250 out of 2250 | elapsed: 1.4s finished
[Parallel(n_jobs=8)]: Done 1784 tasks | elapsed: 1.1s
[Parallel(n_jobs=8)]: Done 2250 out of 2250 | elapsed: 1.4s finished
[Parallel(n_jobs=8)]: Done 2250 out of 2250 | elapsed: 1.4s finished
[Parallel(n_jobs=8)]: Done 2250 out of 2250 | elapsed: 1.2s finished
[Parallel(n_jobs=-1)]: Done 80 out of 80 | elapsed: 9.0min finished
[Parallel(n_jobs=-1)]: Done 34 tasks | elapsed: 0.1s
[Parallel(n_jobs=-1)]: Done 184 tasks | elapsed: 0.5s
[Parallel(n_jobs=-1)]: Done 434 tasks | elapsed: 1.0s
[Parallel(n_jobs=-1)]: Done 784 tasks | elapsed: 1.9s
[Parallel(n_jobs=-1)]: Done 1234 tasks | elapsed: 3.0s
[Parallel(n_jobs=-1)]: Done 1784 tasks | elapsed: 4.3s
```

Search found the best params: {'max_features': 'sqrt', 'n_estimators': 2000}
ROC AUC Train Score: 0.9953199365649816

```
[Parallel(n_jobs=-1)]: Done 2000 out of 2000 | elapsed: 4.8s finished
```

```
rf
```

```
RandomForestClassifier(bootstrap=True, class_weight=None, criterion='gini',
                        max_depth=None, max_features='sqrt', max_leaf_nodes=None,
                        min_impurity_decrease=0.0, min_impurity_split=None,
                        min_samples_leaf=1, min_samples_split=2,
                        min_weight_fraction_leaf=0.0, n_estimators=2000, n_jobs=-1,
                        oob_score=False, random_state=0, verbose=1, warm_start=False)
```

The grid search on this dataset returns the Random Forest model with the same hyperparameter settings as the previous dataset: 2000 trees with subset size $m = \sqrt{p} = 7$.

```
f = plt.figure(figsize=(15,5))
bar_x_axis = range(len(rf.feature_importances_))
plt.bar(bar_x_axis, rf.feature_importances_, width=0.9, color="g",
        tick_label=bar_x_axis)
plt.tick_params(labelsize=10)
plt.show()
```

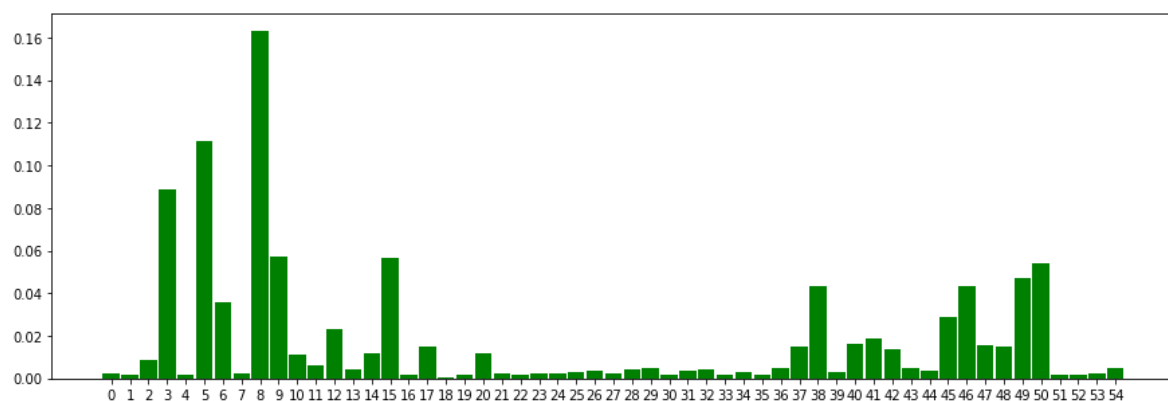


Figure 18: Bar chart of the Random Forest's feature importances agrees with the heatmap and the biplot in that there appears to be 8-10 features that it considers important in explaining most of the information.

Prediction Performance

```
rf_y_pred = rf.predict(x_test)
rf_cm = confusion_matrix(y_true=y_test, y_pred=rf_y_pred)
rf_cm
```

```
[Parallel(n_jobs=8)]: Done 34 tasks | elapsed: 0.0s
[Parallel(n_jobs=8)]: Done 184 tasks | elapsed: 0.0s
[Parallel(n_jobs=8)]: Done 434 tasks | elapsed: 0.1s
[Parallel(n_jobs=8)]: Done 784 tasks | elapsed: 0.1s
[Parallel(n_jobs=8)]: Done 1234 tasks | elapsed: 0.2s
[Parallel(n_jobs=8)]: Done 1784 tasks | elapsed: 0.3s
[Parallel(n_jobs=8)]: Done 2000 out of 2000 | elapsed: 0.3s finished
```

```
array([[372, 14],
       [12, 778]])
```

The Random Forest model has predicted 372 true negatives and 778 true positives.

```
pr_sheep['Random Forest'] = precision_recall(rf_cm)
pr_sheep
```

```
{'Random Forest': (98.23232323232324, 98.48101265822785)}
```

The Random Forest model scored a positive prediction accuracy of 98.23% (precision)

The Random Forest model scored a positive prediction accuracy of 98.26% (precision).

The rate of positive instances that are correctly detected is 98.48% (recall/sensitivity). The difference is small as the classes aren't heavily skewed. The precision/recall is expected to be less useful for this dataset.

```
rf_y_score = rf.predict_proba(x_test)[: , 1]
plot_roc(y_test, rf_y_score)
auc_sheep['Random Forest'] = roc_auc_score(y_true=y_test, y_score=rf_y_score)
print("ROC AUC Test Score: {}".format(auc_sheep['Random Forest']))
plt.show()
```

```
[Parallel(n_jobs=8)]: Done 34 tasks | elapsed: 0.0s
[Parallel(n_jobs=8)]: Done 184 tasks | elapsed: 0.0s
[Parallel(n_jobs=8)]: Done 434 tasks | elapsed: 0.1s
[Parallel(n_jobs=8)]: Done 784 tasks | elapsed: 0.1s
[Parallel(n_jobs=8)]: Done 1234 tasks | elapsed: 0.2s
[Parallel(n_jobs=8)]: Done 1784 tasks | elapsed: 0.3s
[Parallel(n_jobs=8)]: Done 2000 out of 2000 | elapsed: 0.3s finished
```

ROC AUC Test Score: 0.9934347740539122

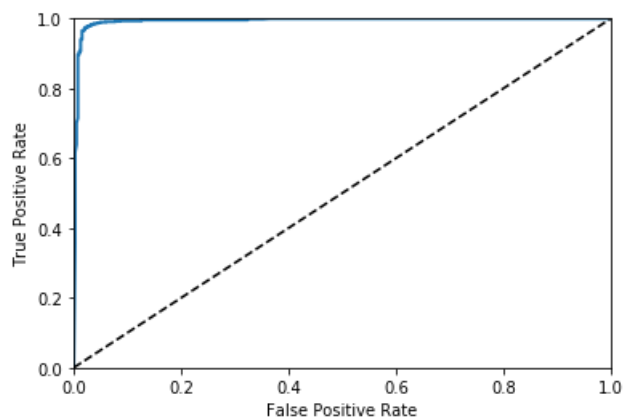


Figure 19: The Random Forest's ROC Curve on the test data.

AdaBoost

```
ab_params = [
    {'n_estimators': [ 800, 1000, 1200, 1400 ], 'learning_rate': [ 0.005, 0.01, 0.05,
0.1 ]}
]
```

```
ab_init = AdaBoostClassifier(random_state=seed)
ab = gridsearchcv_fit(ab_init, ab_params, x_train, y_train)
```

Fitting 10 folds for each of 16 candidates, totalling 160 fits

```
[Parallel(n_jobs=-1)]: Done 34 tasks | elapsed: 2.2min
[Parallel(n_jobs=-1)]: Done 160 out of 160 | elapsed: 9.9min finished
```

Search found the best params: {'learning_rate': 0.05, 'n_estimators': 1200}
ROC AUC Train Score: 0.9940205520042044

ab

```
AdaBoostClassifier(algorithm='SAMME.R', base_estimator=None,
    learning_rate=0.05, n_estimators=1200, random_state=0)
```

This search found the same number as in the MNIST dataset of 1200 trees to be optimal. The learning rate was slightly faster at 0.05 (compared to 0.01) to find optimal performance.

Feature Importances

```
f = plt.figure(figsize=(15,5))
bar_x_axis = range(len(ab.feature_importances_))
plt.bar(bar_x_axis, ab.feature_importances_, width=0.9, color="g",
        tick_label=bar_x_axis)
plt.tick_params(labelsize=10)
plt.show()
```

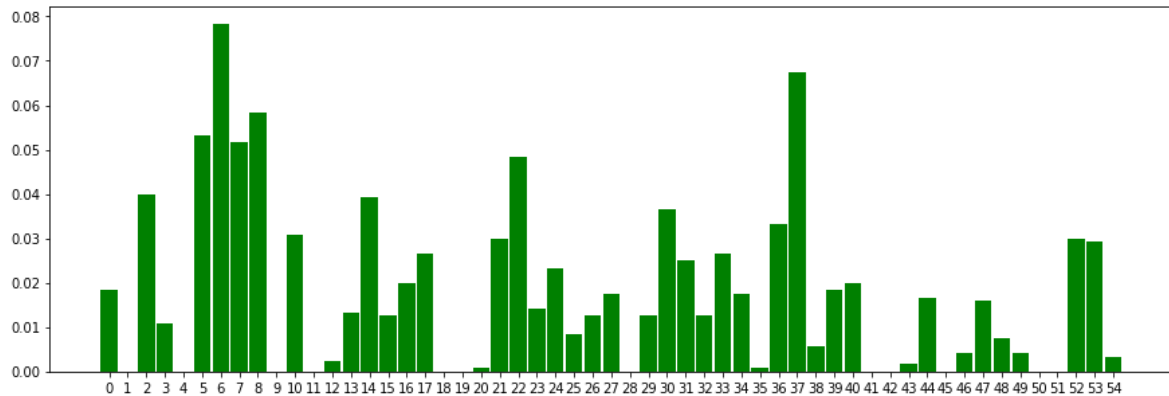


Figure 20: Bar chart of AdaBoost's feature importances. Like the previous dataset, AdaBoost attributes a few a more features with higher importance.

Prediction Performance

```
ab_y_pred = ab.predict(x_test)
ab_cm = confusion_matrix(y_true=y_test, y_pred=ab_y_pred)
ab_cm
```

```
array([[365,  21],
       [ 13, 777]])
```

The AdaBoost model has predicted 365 true negatives and 777 true positives.

```
pr_sheep['AdaBoost'] = precision_recall(ab_cm)
pr_sheep
```

```
{ 'AdaBoost': (97.36842105263158, 98.35443037974684),
  'Random Forest': (98.23232323232324, 98.48101265822785) }
```

The AdaBoost model has a positive prediction accuracy of 97.37% (precision).
The rate of positive instances that are correctly detected is 98.35% (recall/sensitivity).

```
ab_y_score = ab.predict_proba(x_test)[: , 1]
plot_roc(y_test, ab_y_score)
auc_sheep['AdaBoost'] = roc_auc_score(y_true=y_test, y_score=ab_y_score)
print("ROC AUC Test Score: {}".format(auc_sheep['AdaBoost']))
plt.show()
```

ROC AUC Test Score: 0.9914802912048272



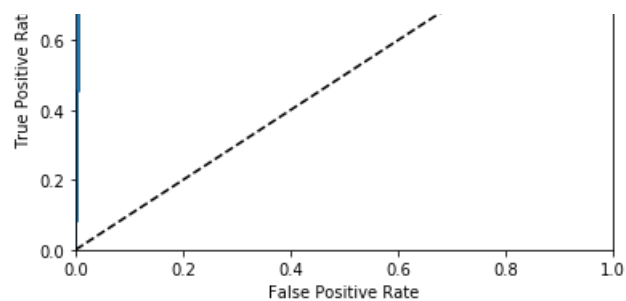


Figure 21: AdaBoost's ROC Curve on the test data.

```
auc_sheep
```

```
{'AdaBoost': 0.9914802912048272, 'Random Forest': 0.9934347740539122}
```

Like the previous dataset, Random Forest comes out on top, almost with the same difference.

XGBoost

```
xg_params = [
    {'n_estimators': [ 1600, 1800, 2000, 2200 ], 'learning_rate': [ 0.0001, 0.001, 0.01
, 0.1 ]},
    {'max_depth': [ 8, 12, 16 ], 'colsample_bytree': [ np.sqrt(p_reduced) / p_reduced, 1
/ 3, 1 ]}]
]
```

```
xg_init = XGBClassifier(random_state=seed)
xg = gridsearchcv_fit(xg_init, xg_params, x_train, y_train)
```

Fitting 10 folds for each of 144 candidates, totalling 1440 fits

```
[Parallel(n_jobs=-1)]: Done 34 tasks      | elapsed: 58.1s
[Parallel(n_jobs=-1)]: Done 184 tasks     | elapsed: 5.0min
[Parallel(n_jobs=-1)]: Done 434 tasks     | elapsed: 10.0min
[Parallel(n_jobs=-1)]: Done 784 tasks     | elapsed: 26.3min
[Parallel(n_jobs=-1)]: Done 1234 tasks    | elapsed: 72.1min
[Parallel(n_jobs=-1)]: Done 1440 out of 1440 | elapsed: 85.2min finished
```

```
Search found the best params: {'max_depth': 12, 'n_estimators': 2000, 'learning_rate': 0.001, 'colsample_bytree':
0.3333333333333333}
ROC AUC Train Score: 0.995364844382323
```

```
xg
```

```
XGBClassifier(base_score=0.5, booster='gbtree', colsample_bylevel=1,
colsample_bytree=0.3333333333333333, gamma=0, learning_rate=0.001,
max_delta_step=0, max_depth=12, min_child_weight=1, missing=None,
n_estimators=2000, n_jobs=1, nthread=None,
objective='binary:logistic', random_state=0, reg_alpha=0,
reg_lambda=1, scale_pos_weight=1, seed=None, silent=True,
subsample=1)
```

This time, the dataset permitted a much higher number of trees (2000 compared to the 1200 in MNIST) before exhibiting signs of overfitting. To settle on the optimal model, a slower learning rate was selected. Similar to the MNIST dataset, injecting random decorrelation via `colsample_bytree` set to 0.33 was found to be significantly useful. Also, the trees used were adjusted to be highly complex.

Feature Importances

```
f = plt.figure(figsize=(15,5))
bar_x_axis = range(len(xg.feature_importances_))
plt.bar(bar_x_axis, xg.feature_importances_, width=0.9, color="g",
tick_label=bar_x_axis)
plt.tick_params(labelsize=10)
plt.show()
```

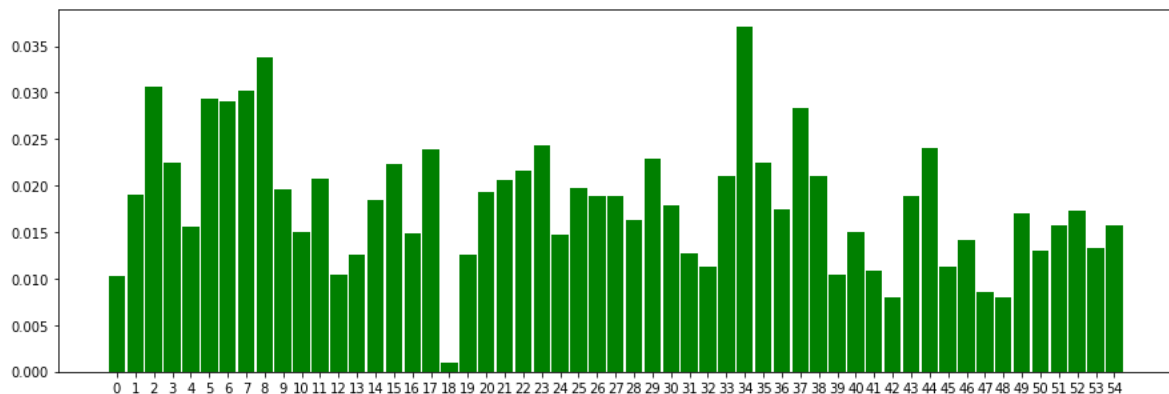


Figure 22: Bar chart of XGBoost's feature importances. Like in the previous dataset, XGBoost distributes the importances more evenly.

Prediction Performance

```
xg_y_pred = xg.predict(x_test)
xg_cm = confusion_matrix(y_true=y_test, y_pred=xg_y_pred)
xg_cm
```

/home/cab/.local/lib/python3.5/site-packages/sklearn/preprocessing/label.py:151: DeprecationWarning: The truth value of an empty array is ambiguous. Returning False, but in future this will result in an error. Use `array.size > 0` to check that an array is not empty.

```
if diff:
```

```
array([[370, 16],
       [ 13, 777]])
```

The XGBoost model has predicted 370 true negatives and 777 true positives.

```
pr_sheep['XGBoost'] = precision_recall(xg_cm)
pr_sheep
```

```
{'AdaBoost': (97.36842105263158, 98.35443037974684),
 'Random Forest': (98.23232323232324, 98.48101265822785),
 'XGBoost': (97.98234552332913, 98.35443037974684)}
```

The XGBoost model has a positive prediction accuracy of 97.98% (precision).
The rate of positive instances that are correctly detected is 98.35% (recall/sensitivity).

```
xg_y_score = xg.predict_proba(x_test)[:, 1]
plot_roc(y_test, xg_y_score)
auc_sheep['XGBoost'] = roc_auc_score(y_true=y_test, y_score=xg_y_score)
print("ROC AUC Test Score: {}".format(auc_sheep['XGBoost']))
plt.show()
```

ROC AUC Test Score: 0.9932232570341706



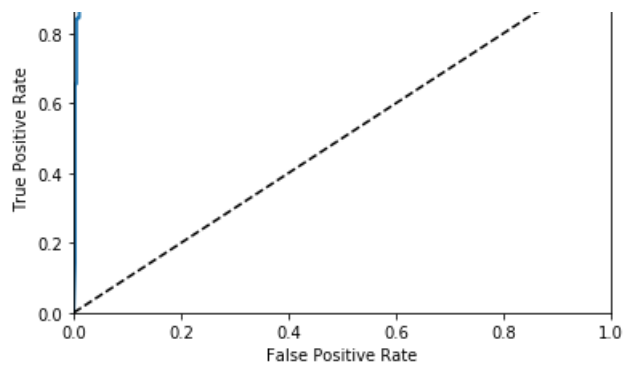


Figure 23: XBoost's ROC Curve on the test data.

auc_sheep

```
{ 'AdaBoost': 0.9914802912048272,
  'Random Forest': 0.9934347740539122,
  'XGBoost': 0.9932232570341706 }
```

This time, the Random Forest model remains on top (whereas the XGBoost was best in the MNIST dataset).

Model Stacking

Meta Learner Data Preparation

```
rf_oobtrain, rf_oobtest = get_oob(rf, x_train, y_train, x_test, seed)
ab_oobtrain, ab_oobtest = get_oob(ab, x_train, y_train, x_test, seed)
xg_oobtrain, xg_oobtest = get_oob(xg, x_train, y_train, x_test, seed)
print("OoB predictions complete!")
```

```
[Parallel(n_jobs=-1)]: Done 34 tasks | elapsed: 0.1s
[Parallel(n_jobs=-1)]: Done 184 tasks | elapsed: 0.4s
[Parallel(n_jobs=-1)]: Done 434 tasks | elapsed: 0.9s
[Parallel(n_jobs=-1)]: Done 784 tasks | elapsed: 1.6s
[Parallel(n_jobs=-1)]: Done 1234 tasks | elapsed: 2.5s
[Parallel(n_jobs=-1)]: Done 1784 tasks | elapsed: 3.6s
[Parallel(n_jobs=-1)]: Done 2000 out of 2000 | elapsed: 4.0s finished
[Parallel(n_jobs=8)]: Done 34 tasks | elapsed: 0.0s
[Parallel(n_jobs=8)]: Done 184 tasks | elapsed: 0.0s
[Parallel(n_jobs=8)]: Done 434 tasks | elapsed: 0.1s
[Parallel(n_jobs=8)]: Done 784 tasks | elapsed: 0.1s
[Parallel(n_jobs=8)]: Done 1234 tasks | elapsed: 0.2s
[Parallel(n_jobs=8)]: Done 1784 tasks | elapsed: 0.3s
[Parallel(n_jobs=8)]: Done 2000 out of 2000 | elapsed: 0.3s finished
[Parallel(n_jobs=8)]: Done 34 tasks | elapsed: 0.0s
[Parallel(n_jobs=8)]: Done 184 tasks | elapsed: 0.0s
[Parallel(n_jobs=8)]: Done 434 tasks | elapsed: 0.1s
[Parallel(n_jobs=8)]: Done 784 tasks | elapsed: 0.1s
[Parallel(n_jobs=8)]: Done 1234 tasks | elapsed: 0.2s
[Parallel(n_jobs=8)]: Done 1784 tasks | elapsed: 0.3s
[Parallel(n_jobs=8)]: Done 2000 out of 2000 | elapsed: 0.3s finished
[Parallel(n_jobs=-1)]: Done 34 tasks | elapsed: 0.1s
[Parallel(n_jobs=-1)]: Done 184 tasks | elapsed: 0.4s
[Parallel(n_jobs=-1)]: Done 434 tasks | elapsed: 0.9s
[Parallel(n_jobs=-1)]: Done 784 tasks | elapsed: 1.6s
[Parallel(n_jobs=-1)]: Done 1234 tasks | elapsed: 2.5s
[Parallel(n_jobs=-1)]: Done 1784 tasks | elapsed: 3.6s
[Parallel(n_jobs=-1)]: Done 2000 out of 2000 | elapsed: 4.0s finished
[Parallel(n_jobs=8)]: Done 34 tasks | elapsed: 0.0s
[Parallel(n_jobs=8)]: Done 184 tasks | elapsed: 0.0s
[Parallel(n_jobs=8)]: Done 434 tasks | elapsed: 0.1s
[Parallel(n_jobs=8)]: Done 784 tasks | elapsed: 0.1s
[Parallel(n_jobs=8)]: Done 1234 tasks | elapsed: 0.2s
[Parallel(n_jobs=8)]: Done 1784 tasks | elapsed: 0.3s
[Parallel(n_jobs=8)]: Done 2000 out of 2000 | elapsed: 0.3s finished
[Parallel(n_jobs=8)]: Done 34 tasks | elapsed: 0.0s
[Parallel(n_jobs=8)]: Done 184 tasks | elapsed: 0.0s
[Parallel(n_jobs=8)]: Done 434 tasks | elapsed: 0.1s
```

[illegible]

[illegible]

[illegible]


```
...
value of an empty array is ambiguous. Returning False, but in future this will result in an error. Use `array.size
> 0` to check that an array is not empty.
    if diff:
```

OoB predictions complete!

```
/home/cab/.local/lib/python3.5/site-packages/sklearn/preprocessing/label.py:151: DeprecationWarning: The truth va
lue of an empty array is ambiguous. Returning False, but in future this will result in an error. Use `array.size
> 0` to check that an array is not empty.
    if diff:
```

```
x_train_meta = np.concatenate((rf_oobtrain, ab_oobtrain, xg_oobtrain), axis=1)
x_test_meta = np.concatenate((rf_oobtest, ab_oobtest, xg_oobtest), axis=1)
print("Meta learner's input data sizes:\nTrain: {}, Test: {}".format(x_train_meta.shape
, x_test_meta.shape))
```

Meta learner's input data sizes:
Train: (4000, 3), Test: (1176, 3)

Simple Majority Vote

```
mv_y_scores = x_test_meta.mean(axis=1).reshape(-1,1)
mv_y_pred = np.around(mv_y_scores)
```

Prediction Performance

```
mv_cm = confusion_matrix(y_true=y_test, y_pred=mv_y_pred)
mv_cm
```

```
array([[371, 15],
       [ 12, 778]])
```

The Majority Vote model has predicted 371 true negatives and 778 true positives.

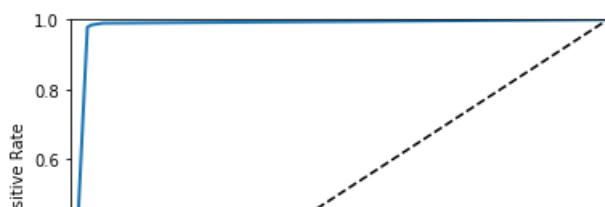
```
pr_sheep['Majority Vote'] = precision_recall(mv_cm)
pr_sheep
```

```
{ 'AdaBoost': (97.36842105263158, 98.35443037974684),
  'Majority Vote': (98.10844892812106, 98.48101265822785),
  'Random Forest': (98.23232323232324, 98.48101265822785),
  'XGBoost': (97.98234552332913, 98.35443037974684) }
```

The Majority Vote model has a positive prediction accuracy of 98.11% (precision).
The rate of positive instances that are correctly detected is 98.48% (recall/sensitivity).

```
plot_roc(y_test, mv_y_scores)
auc_sheep['Majority Vote'] = roc_auc_score(y_true=y_test, y_score=mv_y_scores)
print("ROC AUC Test Score: {}".format(auc_sheep['Majority Vote']))
plt.show()
```

ROC AUC Test Score: 0.9789548763691218



[illegible]

[illegible]

[illegible]

[illegible]

```
[Parallel(n_jobs=-1)]: Done 34 tasks      | elapsed:    0.0s
[Parallel(n_jobs=-1)]: Done 34 tasks      | elapsed:    0.0s
[Parallel(n_jobs=8)]: Done 184 tasks      | elapsed:    0.0s
[Parallel(n_jobs=8)]: Done 34 tasks      | elapsed:    0.0s
[Parallel(n_jobs=8)]: Done 400 out of 400 | elapsed:    0.1s finished
[Parallel(n_jobs=8)]: Done 184 tasks      | elapsed:    0.1s
[Parallel(n_jobs=8)]: Done 34 tasks      | elapsed:    0.0s
[Parallel(n_jobs=-1)]: Done 184 tasks      | elapsed:    0.1s
[Parallel(n_jobs=-1)]: Done 184 tasks      | elapsed:    0.1s
[Parallel(n_jobs=8)]: Done 184 tasks      | elapsed:    0.1s
[Parallel(n_jobs=8)]: Done 400 out of 400 | elapsed:    0.1s finished
[Parallel(n_jobs=8)]: Done 400 out of 400 | elapsed:    0.1s finished
[Parallel(n_jobs=-1)]: Done 400 out of 400 | elapsed:    0.3s finished
[Parallel(n_jobs=-1)]: Done 400 out of 400 | elapsed:    0.3s finished
[Parallel(n_jobs=8)]: Done 34 tasks      | elapsed:    0.0s
[Parallel(n_jobs=8)]: Done 184 tasks      | elapsed:    0.0s
[Parallel(n_jobs=8)]: Done 400 out of 400 | elapsed:    0.1s finished
[Parallel(n_jobs=8)]: Done 34 tasks      | elapsed:    0.0s
[Parallel(n_jobs=8)]: Done 184 tasks      | elapsed:    0.0s
[Parallel(n_jobs=8)]: Done 34 tasks      | elapsed:    0.0s
[Parallel(n_jobs=8)]: Done 400 out of 400 | elapsed:    0.1s finished
[Parallel(n_jobs=8)]: Done 184 tasks      | elapsed:    0.0s
[Parallel(n_jobs=8)]: Done 34 tasks      | elapsed:    0.0s
[Parallel(n_jobs=8)]: Done 400 out of 400 | elapsed:    0.1s finished
[Parallel(n_jobs=8)]: Done 184 tasks      | elapsed:    0.0s
[Parallel(n_jobs=8)]: Done 400 out of 400 | elapsed:    0.1s finished
```

```
Search found the best params: {'n_estimators': 20}
ROC AUC Train Score: 0.9762698470332026
```

```
[Parallel(n_jobs=-1)]: Done 60 out of 60 | elapsed:    5.0s finished
[Parallel(n_jobs=-1)]: Done 20 out of 20 | elapsed:    0.0s finished
```

```
metarf
```

```
RandomForestClassifier(bootstrap=True, class_weight=None, criterion='gini',
                        max_depth=None, max_features='sqrt', max_leaf_nodes=None,
                        min_impurity_decrease=0.0, min_impurity_split=None,
                        min_samples_leaf=1, min_samples_split=2,
                        min_weight_fraction_leaf=0.0, n_estimators=20, n_jobs=-1,
                        oob_score=False, random_state=0, verbose=1, warm_start=False)
```

This time, the Meta-Forest grew slightly larger to 20 trees before overfitting.

Prediction Performance

```
metarf_y_pred = metarf.predict(x_test_meta)
metarf_cm = confusion_matrix(y_true=y_test, y_pred=metarf_y_pred)
metarf_cm
```

```
[Parallel(n_jobs=8)]: Done 20 out of 20 | elapsed:    0.0s finished
```

```
array([[373, 13],
       [13, 777]])
```

The Meta-Forest model has predicted 373 true negatives and 777 true positives.

```
pr_sheep['Meta-Forest'] = precision_recall(metarf_cm)
pr_sheep
```

```
{ 'AdaBoost': (97.36842105263158, 98.35443037974684),
  'Majority Vote': (98.10844892812106, 98.48101265822785),
  'Meta-Forest': (98.35443037974684, 98.35443037974684),
  'Random Forest': (98.23232323232324, 98.48101265822785),
  'XGBoost': (97.98234552332913, 98.35443037974684) }
```

The Meta-Forest model has a positive prediction accuracy of 98.35% (precision).
The rate of positive instances that are correctly detected is 98.35% (recall/sensitivity).

```
metarf_y_score = metarf.predict_proba(x_test_meta)[: , 1]
plot_roc(y_test, metarf_y_score)
auc_sheep['Meta-Forest'] = roc_auc_score(y_true=y_test, y_score=metarf_y_score)
print("ROC AUC Test Score: {}".format(auc_sheep['Meta-Forest']))
plt.show()
```

[Parallel(n_jobs=8)]: Done 20 out of 20 | elapsed: 0.0s finished

ROC AUC Test Score: 0.9789515970354824

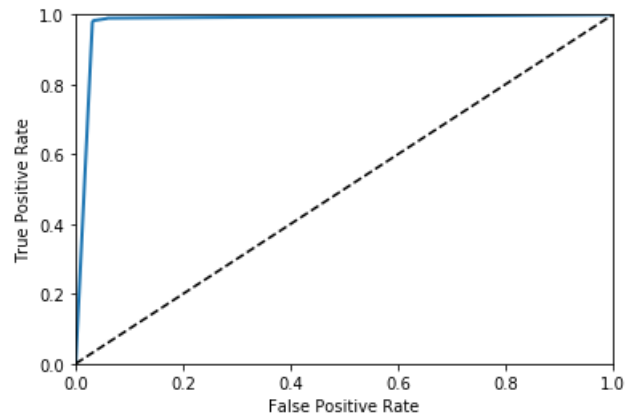


Figure 25: ROC Curve on the test data from the Meta-Forest.

auc_sheep

```
{'AdaBoost': 0.9914802912048272,
 'Majority Vote': 0.9789548763691218,
 'Meta-Forest': 0.9789515970354824,
 'Random Forest': 0.9934347740539122,
 'XGBoost': 0.9932232570341706}
```

Meta Learner XGBoost

```
metaxg_params = [
    {'n_estimators': [ 800, 1000, 1200, 1400, 1400 ], 'learning_rate': [ 1e-12, 1e-10, 1
e-08, 1e-06 ],
    'max_depth': [ 1, 2, 4 ], 'colsample_bytree': [ 1 / 3, 2 / 3, 1 ] }
]
```

```
metaxg_init = XGBClassifier(random_state=seed)
metaxg = gridsearchcv_fit(metaxg_init, metaxg_params, x_train_meta, y_train)
```

Fitting 10 folds for each of 180 candidates, totalling 1800 fits

```
[Parallel(n_jobs=-1)]: Done 34 tasks      | elapsed: 2.3s
[Parallel(n_jobs=-1)]: Done 184 tasks     | elapsed: 12.3s
[Parallel(n_jobs=-1)]: Done 434 tasks     | elapsed: 30.1s
[Parallel(n_jobs=-1)]: Done 784 tasks     | elapsed: 57.6s
[Parallel(n_jobs=-1)]: Done 1234 tasks    | elapsed: 1.6min
[Parallel(n_jobs=-1)]: Done 1784 tasks    | elapsed: 2.6min
[Parallel(n_jobs=-1)]: Done 1800 out of 1800 | elapsed: 2.7min finished
```

```
Search found the best params: {'max_depth': 2, 'n_estimators': 1000, 'learning_rate': 1e-08, 'colsample_bytree':
0.6666666666666666}
```


ROC AUC Train Score: 0.9762644087748122

metaxg

```
XGBClassifier(base_score=0.5, booster='gbtree', colsample_bylevel=1,
              colsample_bytree=0.6666666666666666, gamma=0, learning_rate=1e-08,
              max_delta_step=0, max_depth=2, min_child_weight=1, missing=None,
              n_estimators=1000, n_jobs=1, nthread=None,
              objective='binary:logistic', random_state=0, reg_alpha=0,
              reg_lambda=1, scale_pos_weight=1, seed=None, silent=True,
              subsample=1)
```

This time, the Meta-XGBoost learner found benefit in sampling twice as many features per tree fit. Further, overfitting was encountered sooner, restraining growth to 1000 trees (compared to 1800+ trees in the previous dataset).

Prediction Performance

```
metaxg_y_pred = metaxg.predict(x_test_meta)
metaxg_cm = confusion_matrix(y_true=y_test, y_pred=metaxg_y_pred)
metaxg_cm
```

```
/home/cab/.local/lib/python3.5/site-packages/sklearn/preprocessing/label.py:151: DeprecationWarning: The truth value of an empty array is ambiguous. Returning False, but in future this will result in an error. Use `array.size > 0` to check that an array is not empty.
  if diff:
```

```
array([[371,  15],
       [ 12, 778]])
```

The Meta-XGBoost model has predicted 371 true negatives and 778 true positives.

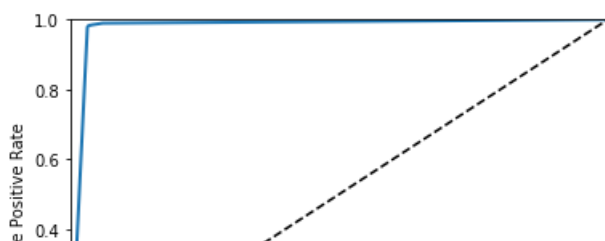
```
pr_sheep['Meta-XGBoost'] = precision_recall(metaxg_cm)
pr_sheep
```

```
{ 'AdaBoost': (97.36842105263158, 98.35443037974684),
  'Majority Vote': (98.10844892812106, 98.48101265822785),
  'Meta-Forest': (98.35443037974684, 98.35443037974684),
  'Meta-XGBoost': (98.10844892812106, 98.48101265822785),
  'Random Forest': (98.23232323232324, 98.48101265822785),
  'XGBoost': (97.98234552332913, 98.35443037974684) }
```

The Meta XGBoost model has a positive prediction accuracy of 98.11% (precision).
The rate of positive instances that are correctly detected is 98.48% (recall/sensitivity).

```
metaxg_y_score = metaxg.predict_proba(x_test_meta)[: , 1]
plot_roc(y_test, metaxg_y_score)
auc_sheep['Meta-XGBoost'] = roc_auc_score(y_true=y_test, y_score=metaxg_y_score)
print("ROC AUC Test Score: {}".format(auc_sheep['Meta-XGBoost']))
plt.show()
```

ROC AUC Test Score: 0.97896471437004



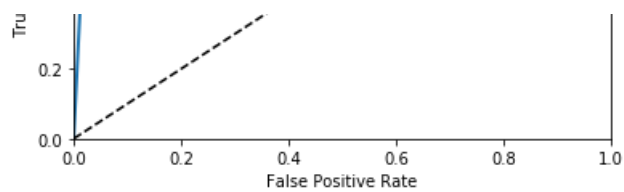


Figure 26: ROC Curve on the test data from the Meta Learner.

```
auc_sheep
```

```
{'AdaBoost': 0.9914802912048272,
 'Majority Vote': 0.9789548763691218,
 'Meta-Forest': 0.9789515970354824,
 'Meta-XGBoost': 0.97896471437004,
 'Random Forest': 0.9934347740539122,
 'XGBoost': 0.9932232570341706}
```

```
auc_sheep = {'AdaBoost': 0.9914802912048272,
 'Majority Vote': 0.9789548763691218,
 'Meta-Forest': 0.9789515970354824,
 'Meta-XGBoost': 0.97896471437004,
 'Random Forest': 0.9934347740539122,
 'XGBoost': 0.9932232570341706}
```

```
auc_mnist = {'AdaBoost': 0.9964689831372104,
 'Majority Vote': 0.9797359353447191,
 'Meta-Forest': 0.9794651908780632,
 'Meta-XGBoost': 0.9797359353447191,
 'Random Forest': 0.9981555217541017,
 'XGBoost': 0.9982809452784596}
```

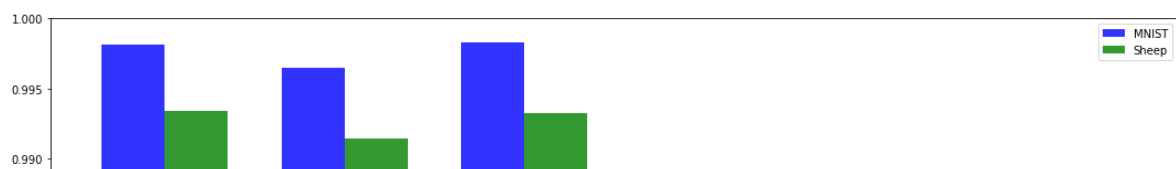
Final Plots and Comments

```
n_groups = len(auc_mnist)

list_mnist = list(range(n_groups))
list_mnist[0] = auc_mnist['Random Forest']
list_mnist[1] = auc_mnist['AdaBoost']
list_mnist[2] = auc_mnist['XGBoost']
list_mnist[3] = auc_mnist['Majority Vote']
list_mnist[4] = auc_mnist['Meta-Forest']
list_mnist[5] = auc_mnist['Meta-XGBoost']

list_sheep = list(range(n_groups))
list_sheep[0] = auc_sheep['Random Forest']
list_sheep[1] = auc_sheep['AdaBoost']
list_sheep[2] = auc_sheep['XGBoost']
list_sheep[3] = auc_sheep['Majority Vote']
list_sheep[4] = auc_sheep['Meta-Forest']
list_sheep[5] = auc_sheep['Meta-XGBoost']
```

```
barchart_double(0.975, 1.0, n_groups, list_mnist, 'MNIST', list_sheep, 'Sheep')
```



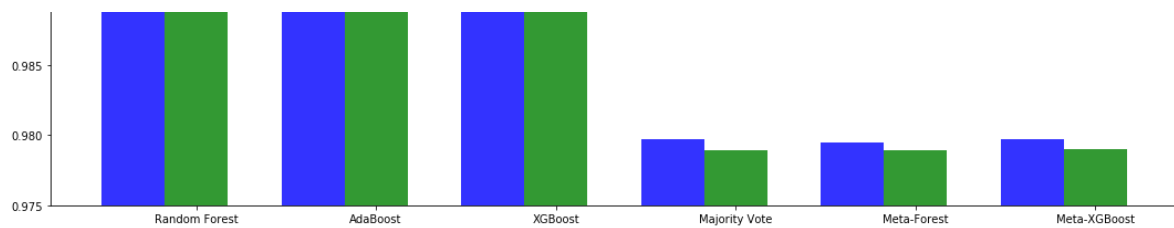


Figure 27: Bar chart of the each model's ROC AUC test scores. First, stacking has proven to fail compared to the non-stacking models. Although still achieving a high score, it is expected that the low number of features (lower layer models) contributed to its difficulty in deciding which predictions to favour. Further, it could be argue that each lower layer model did not provide adequate diversity, which requires either different models, or more models, to improve. Under these circumstances, the naive Majority Vote method remained competitive. Overall XGBoost performed the best, which was somewhat expected given the shared characteristics with both the Random Forest and AdaBoost models in that it was tuned for feature subsetting. However, training were vastly longer to achieve minute advantages over the Random Forest. The degree of parallelism allowed by the Random Forest method cut training times to approximately 3-5% that of XGBoost. This would undoubtedly be a major consideration in production.

As both datasets consisted of highly correlated data, AdaBoost performed the worst out of the non-stacking methods. Even with the increased tree complexity (`max_depth > 1`) present in the XGBoost, Random Forest managed to take close second places in both datasets. These results support the importance of decorrelating the weak learners. In the MNIST analysis, PCA components were used which are inherently correlated where the first greatly dominates the others. Ignoring this strong feature in some of the trees would give a vastly different perspective of the data than in trees that acknowledge this feature. This forced tree diversity is likely to have enhanced the averaging power over all trees, resulting in better noise handling.

```
mnist_precision = list(range(n_groups))
mnist_precision[0] = pr_mnist['Random Forest'][0]
mnist_precision[1] = pr_mnist['AdaBoost'][0]
mnist_precision[2] = pr_mnist['XGBoost'][0]
mnist_precision[3] = pr_mnist['Majority Vote'][0]
mnist_precision[4] = pr_mnist['Meta-Forest'][0]
mnist_precision[5] = pr_mnist['Meta-XGBoost'][0]

mnist_recall = list(range(n_groups))
mnist_recall[0] = pr_mnist['Random Forest'][1]
mnist_recall[1] = pr_mnist['AdaBoost'][1]
mnist_recall[2] = pr_mnist['XGBoost'][1]
mnist_recall[3] = pr_mnist['Majority Vote'][1]
mnist_recall[4] = pr_mnist['Meta-Forest'][1]
mnist_recall[5] = pr_mnist['Meta-XGBoost'][1]
```

```
barchart_double(93, 99, n_groups, mnist_precision, 'Precision', mnist_recall, 'Recall')
```

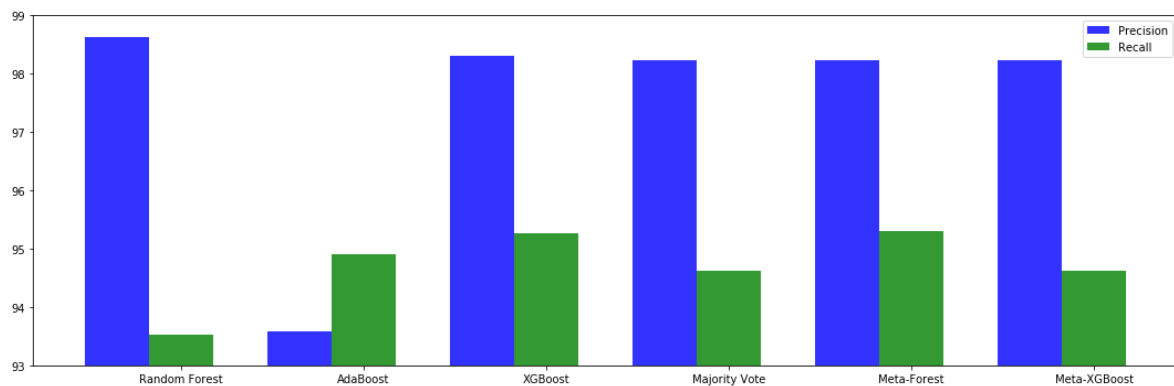


Figure 28: Bar chart of each model's precision and recall score on the heavily skewed MNIST dataset. Interestingly, AdaBoost performs better at recall (rate of correct True Positive detections) than precision (True Positive prediction accuracy). All other models perform better at precision, where Random Forest performs best but all perform roughly the same. The greater differences lie in each of the model's recall performance, where the XGBoost and Meta-Forest perform the best. These results show that in scenarios where either precision or recall performance is favoured rather than overall performance, model selection can differ. If the application's context requires that the model must maintain a high recall performance to sacrifice precision. AdaBoost or the slow

require that the model must maintain a high recall performance to address problem, therefore, the given XGBoost might warrant consideration over the Random Forest.