Exercise 2 – Omri Shlomy

ID 208394718

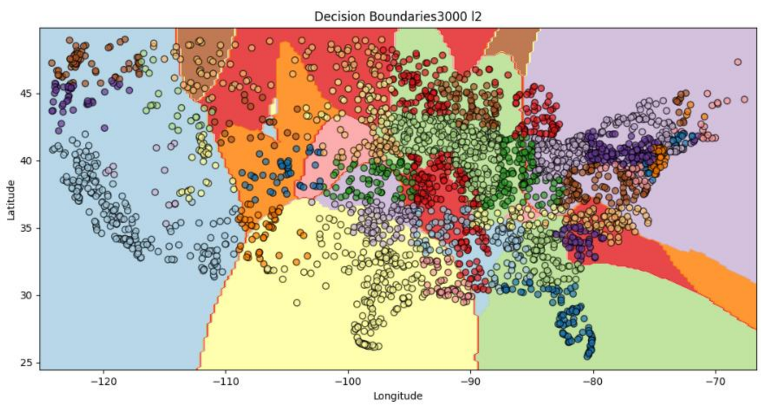
KNN

5.1.3

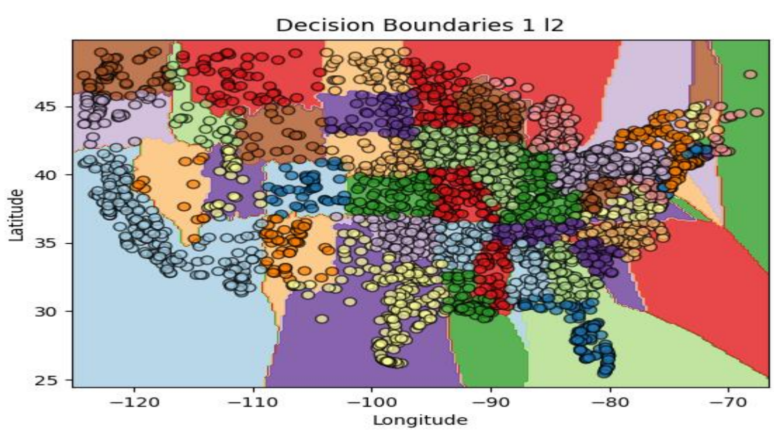
	K=1	K=10	K=100	K=1000	K=3000
L1	0.9670	0.9617	0.9231	0.7450	0.4017
L2	0.9680	0.9577	0.9201	0.7416	0.3981

5.2.2

Kmin with L2



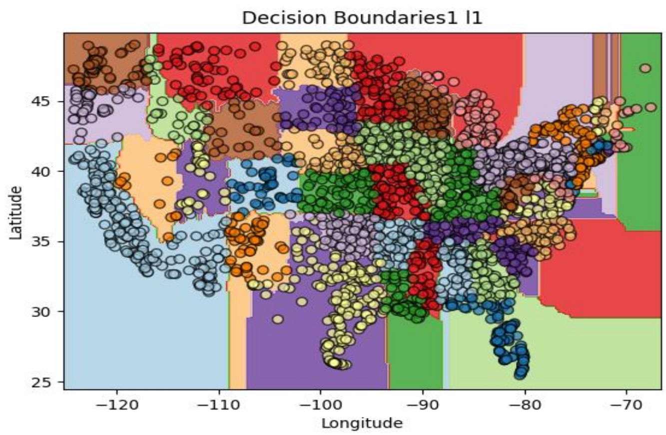
Kmax with L2



(a)

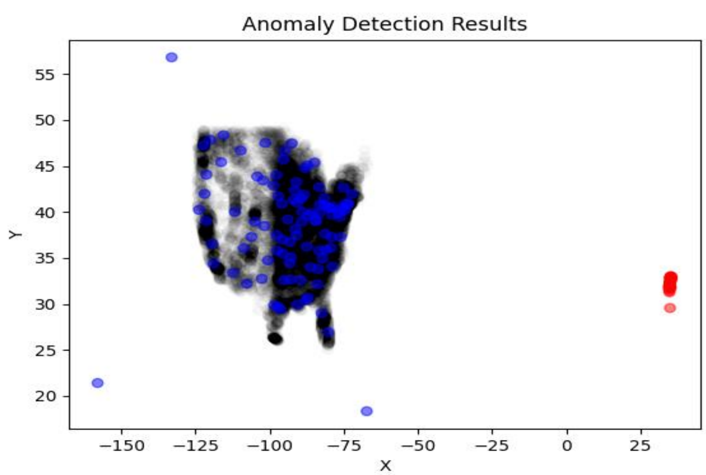
Kmax(1) shows better results then Kmin(3000) because in a higher probability if we look at the closest neighbor to the sample we want to classify, it will be in the same state. However, if we look at 3000 KNN's, in high probability there will be neighbors outside the state – the number of KNN's is not very small from the data set. In this case, most of the neighbors will probably be in another state of the sample's.

(b)



When comparing between l1, which is the absolute between 2 points, with l2, which is the Euclidean distance between points, we can see that the classification of l1 Is more linear with strict lines dividing between states, while l2 is parabolic and has more curve line classifying the data between states.

5.3

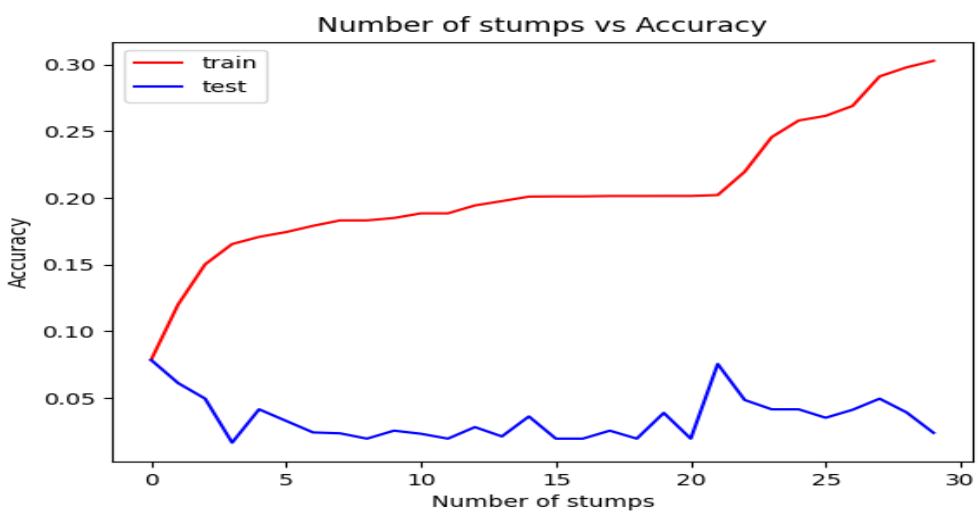


5.4

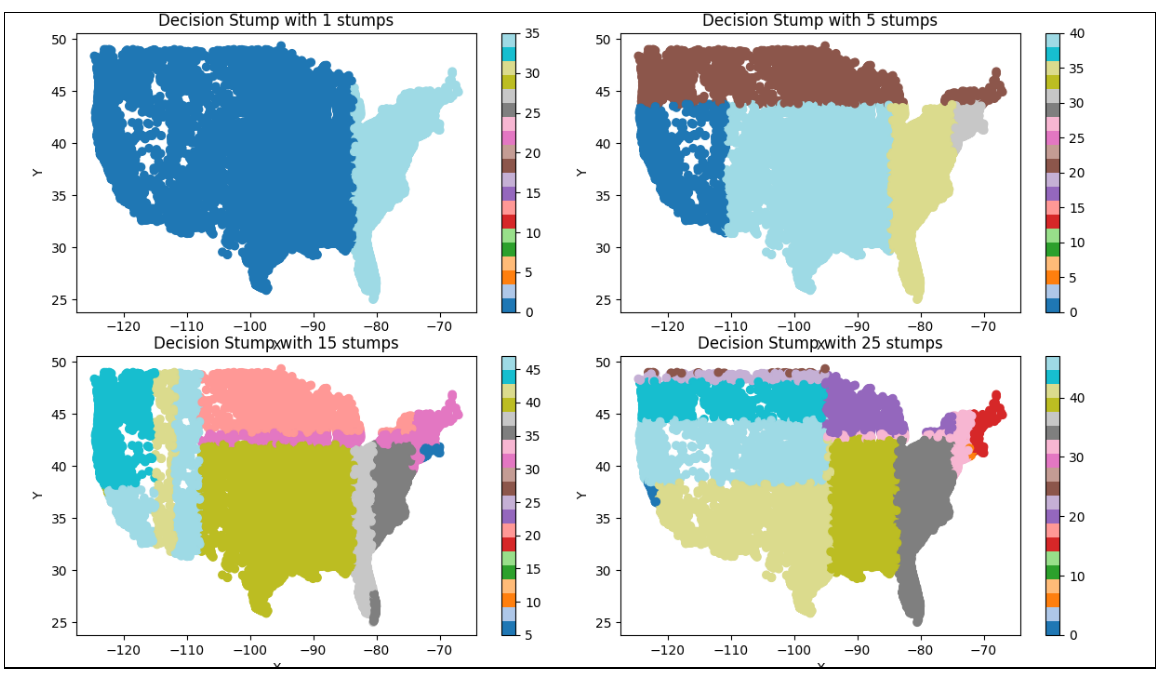
The anomalies in red differ from the data in x axis and are far from mean of all samples. It suggests that these points are located in another area then the US states(Africa or something). There are also 3 additional points that differ from the data distribution and are also located out of the US, implies that the correct numbers of anomalies for this data set is 53.

Greedy boosting

1.3.1



1.3.2



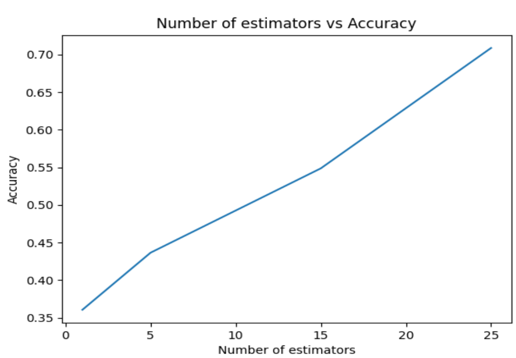
1.3.3

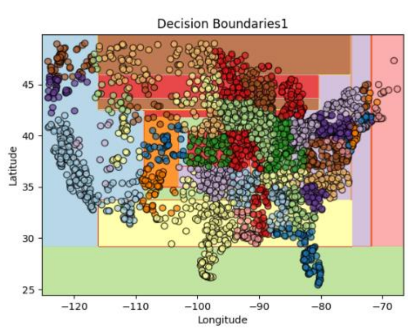
The way greedy boosting learns can be abstract as putting a table over the data set, while each line in the table fits to a chosen threshold. However, the states in the US are far from being a rectangle so they cannot be separated by rectangles. Also, the decision stump trains over and over on the same data and improves feature based on the previous stump,making an overfitting effect on the training data. Even if we could design a stump based weak learner we'd never get to 100% accuracy, because of the shapes of the states that represents the data distribution.

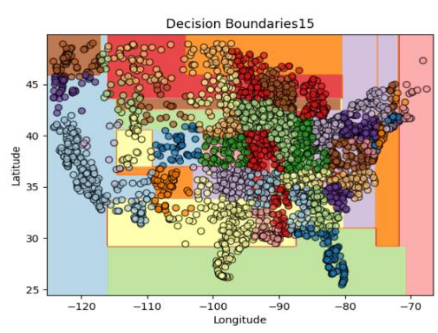
1.3.4

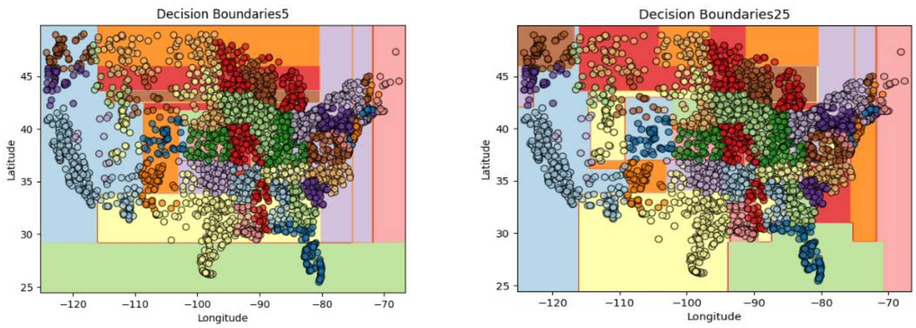
The results of KNN's are much better than the results of greedy boosting,because it describes better the data of the real word. the stump decision thresholds try to govern the data into small rectangle, while the shape of the geographical states are not that way. The KNN's use only the approximation of a sample to samples in it's data set, which represents better the real world.

1.4









we see the predictions in XGB has more classes even with a small number of estimators. This observation can be related to the gradient learning of the XGB compare to the sequential learning of greedy boosting. We also can tell the test accuracy is much better for the XGB even with a small number of estimators, and the learning time is faster.