SMAI (CSE 471)

Spring-2019

Assignment-3

Name: Karnati Venkata Kartheek

Roll Number 2018801010

**Question1**

**Part2(Q1.2)**

KMeans Clustering

purity of cluster 0 is 0.9929 and its dominant class label is dos

purity of cluster 1 is 0.8977 and its dominant class label is normal

purity of cluster 2 is 0.9773 and its dominant class label is dos

purity of cluster 3 is 0.4604 and its dominant class label is dos

purity of cluster 4 is 0.8070 and its dominant class label is probe

**Part3(Q1.3)**

GMM

purity of cluster 0 is 0.7269 and its dominant class label is normal

purity of cluster 1 is 1.0 and its dominant class label is dos

purity of cluster 2 is 0.8455 and its dominant class label is normal

purity of cluster 3 is 0.4213 and its dominant class label is dos

purity of cluster 4 is 0.8852 and its dominant class label is normal

**Part4(Q1.4)**

Hierarchical (AC)

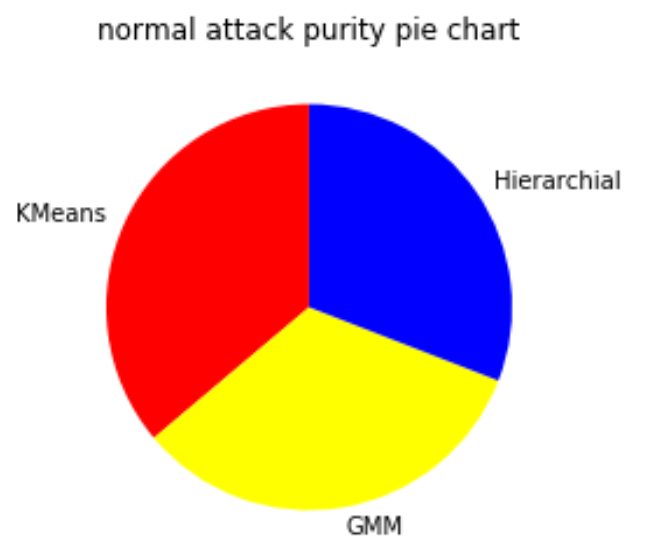
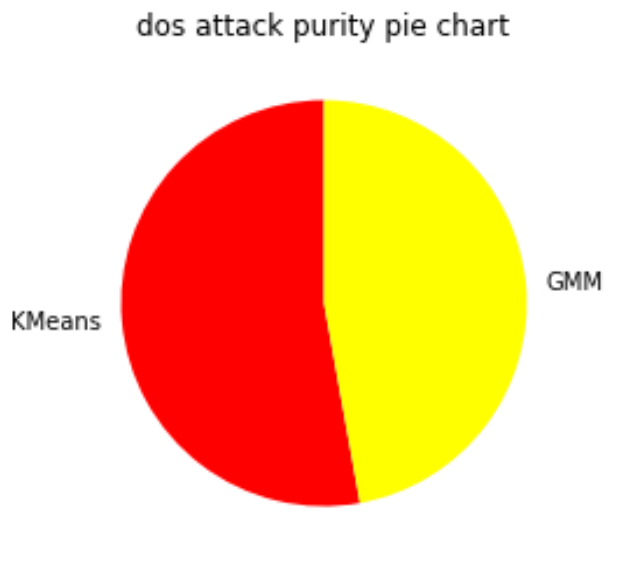
purity of cluster 0 is 0.5348 and its dominant class label is normal

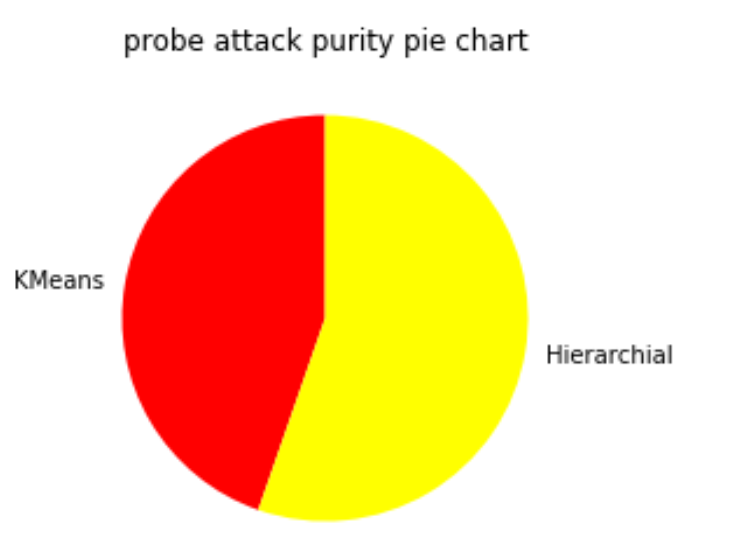
purity of cluster 1 is 1.0 and its dominant class label is normal

purity of cluster 2 is 0.6666 and its dominant class label is r2l

purity of cluster 3 is 1.0 and its dominant class label is probe

purity of cluster 4 is 0.9285 and its dominant class label is r2l



Above pie charts compares the purtiy of 3classes obtained using different clustering methods

**Part5(Q1.5)**

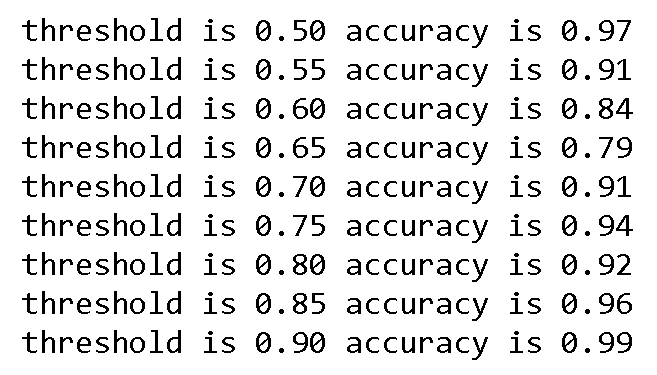
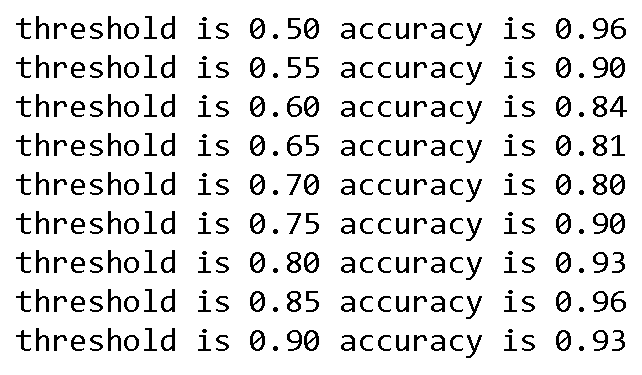
If the data has both categorical and numerical attributes then we convert each categorical feature with k possible outcome to k numerical feature where all are zeros except the one corresponding to the value of categorical attribute and its value is one. Thus we can perform PCA on both categorical and numerical attributes.

**Question2**

**Part2(Q2.2)**

Accuracy values for different threshold values

Logistic regression KNN

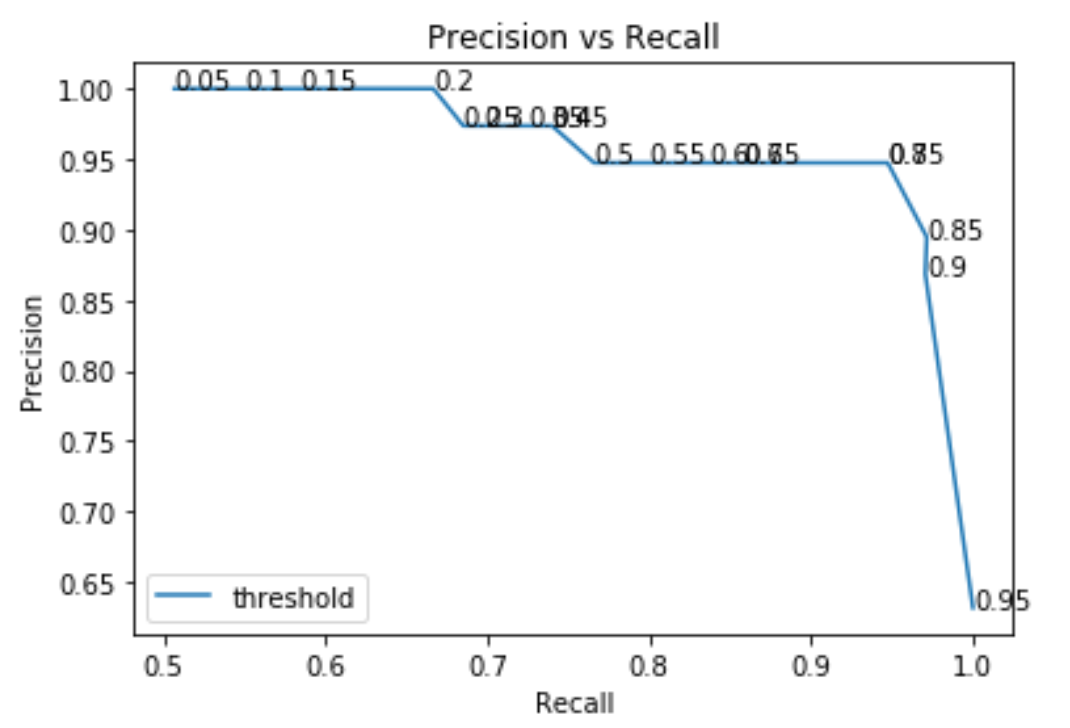
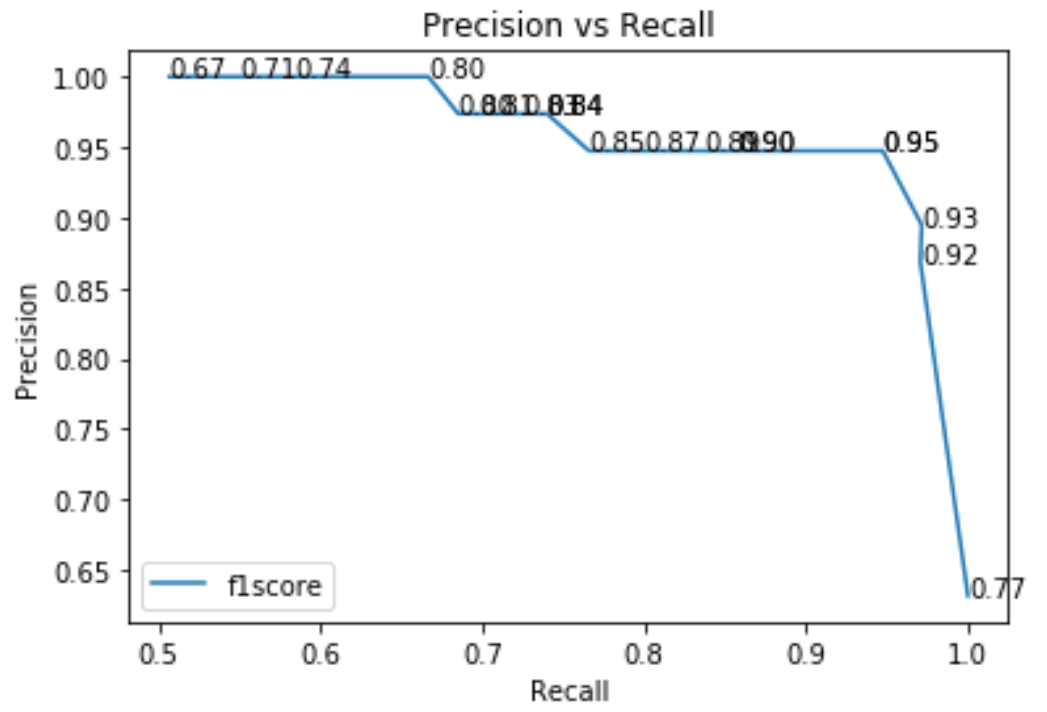
We can see both logistic regression and KNN are performing similarly

**Part3(Q2.3)**

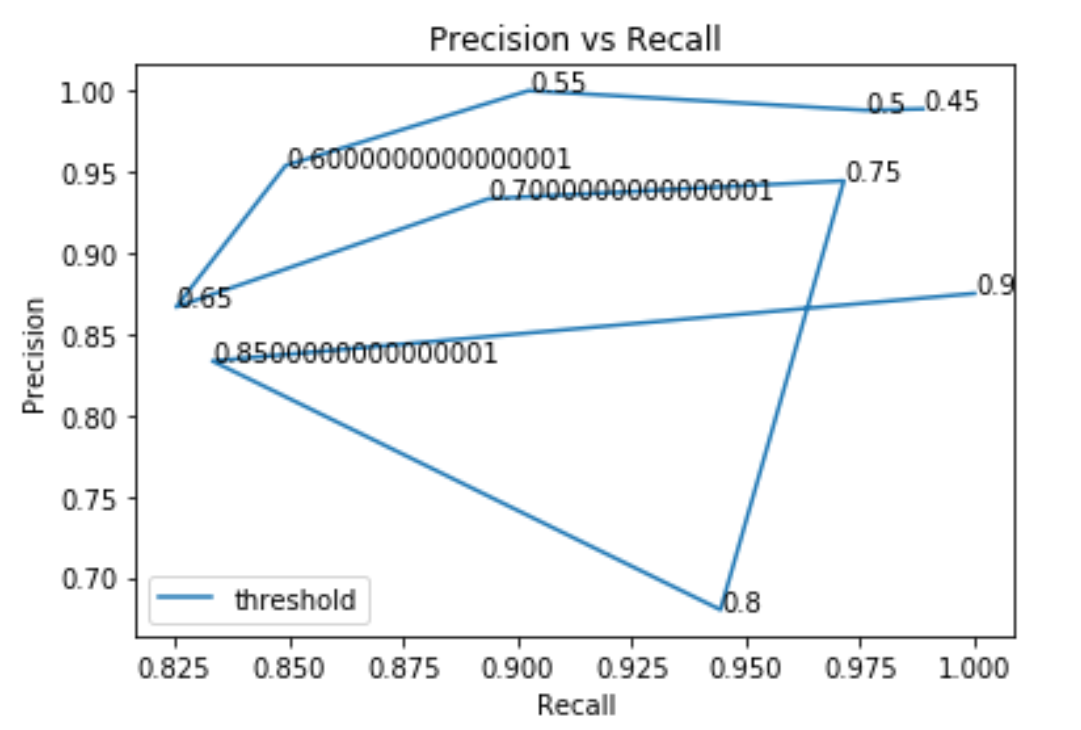
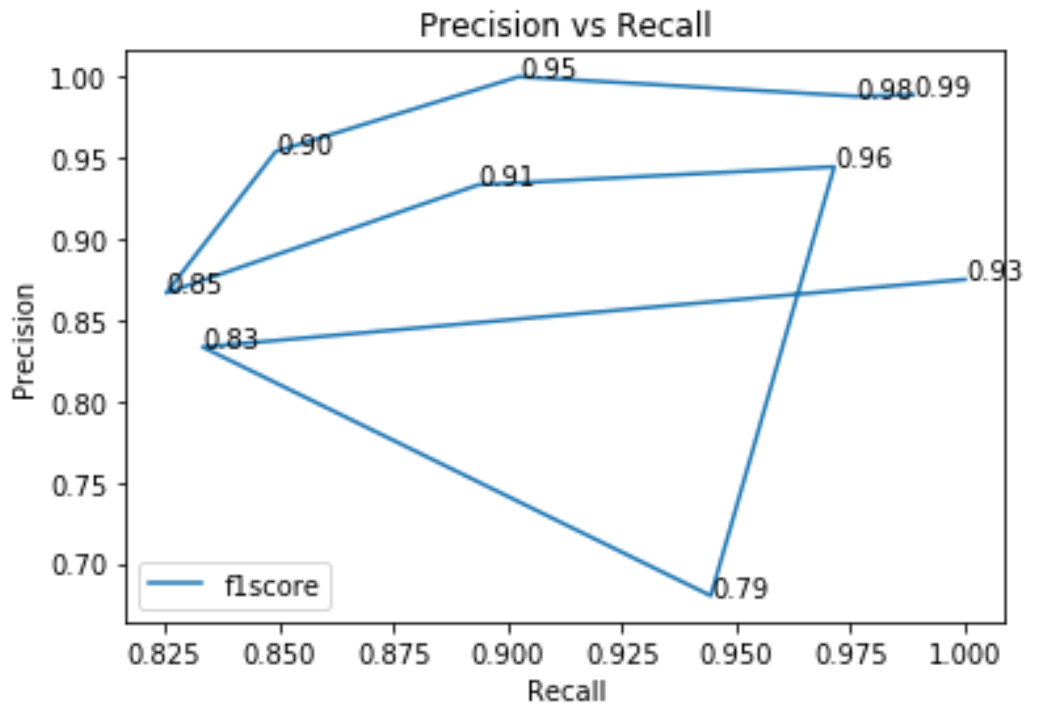
Precision vs Recall (threshold for dividing the classes initially is 0.725)

First graph shows precision vs recall with corresponding threshold in range (0 to 1)

Second graph shows precision vs recall with f1score of corresponding threshold

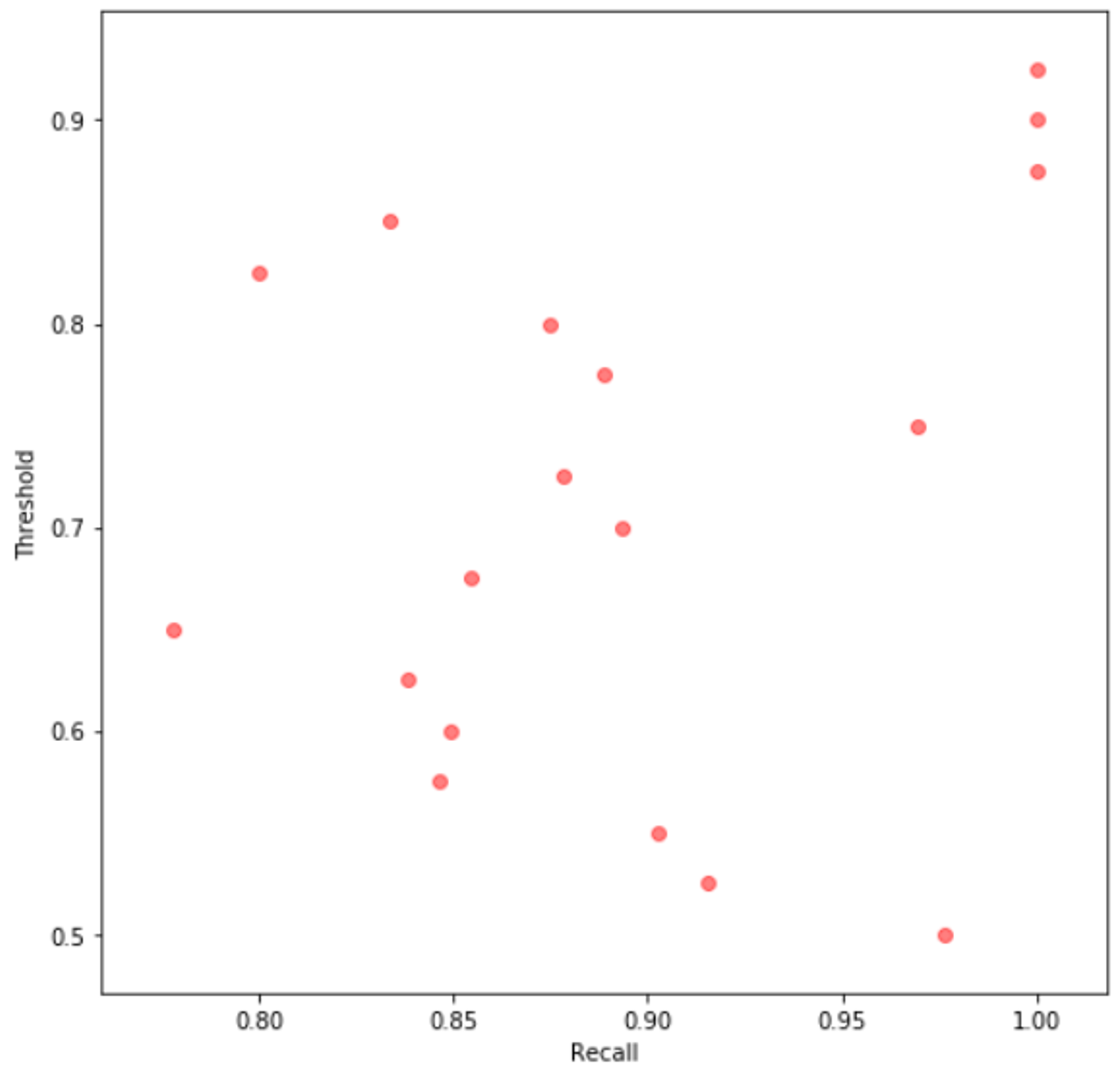
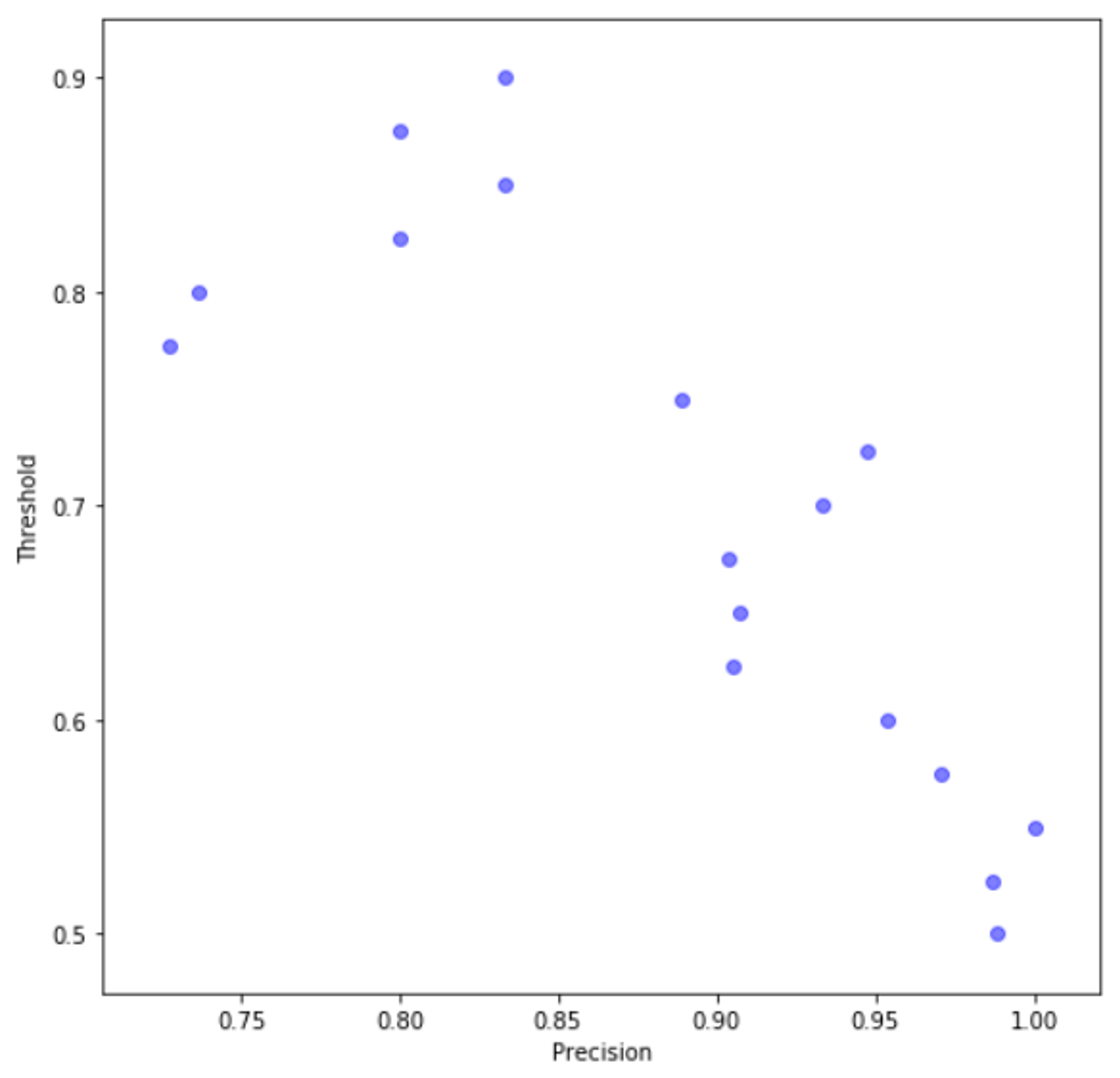
 

Graphs with same threshold for dividing the classes and comparing at time of decsion

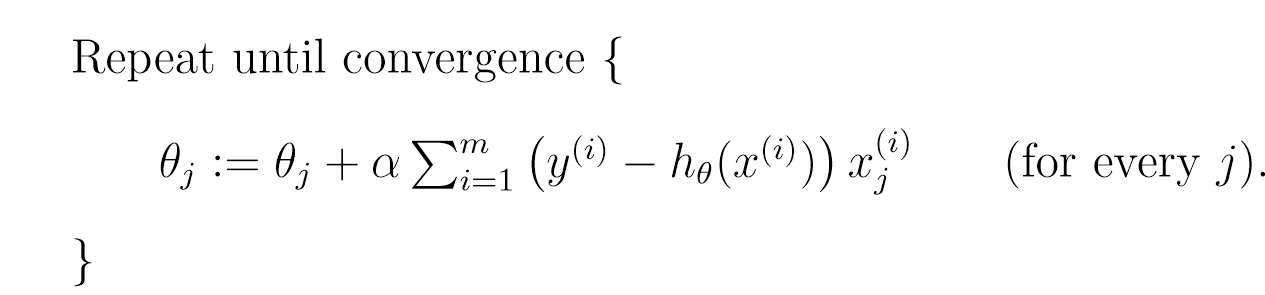
The graphs below shops Threshold vs Precision and recall

Positive corelation with Precision and Negative corelation with recall

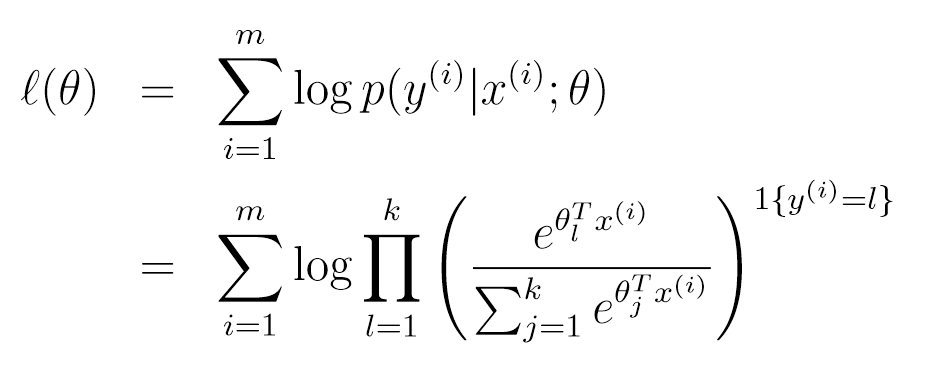
 

**Question3**

Gradient Ascent formula used in One vs All case



Likelihood formula used for deriving gradient ascent equation in One vs One case



Accuracy in One vs All case 0.7182539682539684

Accuracy in One vs One case 0.48299319727891155