Page No.
Date:

Assignment

(1) Small k with noisy data
(2) They always overfit
(3) Reducing variance
(4) improving feature scaling
(5) Errors are normally distributed
(6) Sigmoid
(7) Recall
(8) Training time
(9) To improve accuracy always
(10) KNN

(11) Overfitting in decision tree when using depth as a parameter. is when we use depth there are multiple nodes are showing. The tree will starting the memorized unnecessary data of nodes that is not required.
Bagging will take the data from sample parameter and train it on diffrent ~~param~~ models in a proper way then collect the trained data from models and combine it. It helps the data to be train and to show it its accuracy easily.
Random forest also do the same but in this the model is not specified it train the data on multiple or random models.

(12) Random forest is used to train the dataset from 4 random machine learning models. The best fit model's trained dataset is selected.

a) In bootstrap sampling the each tree looks similer data

b) In random feature selection the randomly trained data is used

b) Majority voting :- Most common or most repeated dataset considered

(13)

Actual grave	120	30
Actual not grave	50	800

Accuracy :- $\frac{120+30}{120} = \frac{150}{120} = 1.2$

TP 120	FN 30
FP 50	TN 800

Precision Accuracy :- $\frac{T.P}{TP+FN} = \frac{120}{120+30} = \frac{120}{150} = 0.8$

$= \frac{}{}$

Precision Recall :- $\frac{TP}{FP+FN} = \frac{120}{50+30} = \frac{120}{80} = 1.5$

$F_1 = \frac{2 \times P \times R}{P+R} = \frac{2 \times 0.8 \times 1.5}{0.8+1.5} = 1.8$

a) Accuracy :- 1.2

b) Precision :- 0.8

c) Recall :- 1.5

d) $F_1 = 1.8$

e) No

(14) a) All features considered as each split

b) This model overfit less overfit than a single decision tree because it doesn't train one by one like decision tree train nodes one by one

c) There is no impact on bias and variance on estimator = 200 because random forest has low bias and low variance feature. But it takes training time