

Q1 Small K with noisy data

Q2 Small changes in data lead to different trees.

Q3 Reducing variance

Q4 All features are considered at each split.

Q5 Features are ~~not~~ independent

Q6 Sigmoid function

Q7 Accuracy.

Q8 Variance

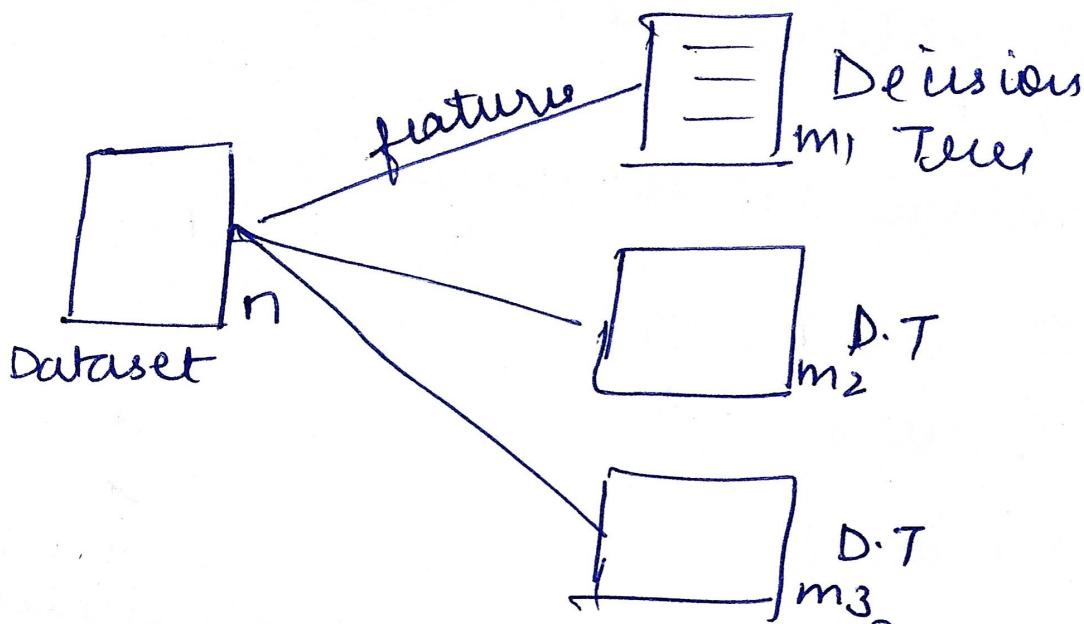
Q9 Because distance calculation depends on scale

Q10 ~~Decision Tree~~ KNN

Q11 Overfitting means when model performs well on training data but not on test data. Here the model learns the data present in the dataset instead of learning pattern.

~~Decision Trees can overfit.~~ As the size of tree ~~increased~~ becomes large & due to addition of more features its depth also increases. As the depth increases ~~the~~ features repeat due to this instead of learning pattern decision tree starts to learn the data present. Thus leads to overfitting.

Overfitting in decision tree is removed or can be solved through Bagging & creation of Random Forest (Ensemble technique).



In bagging a dataset sends its features (with or without replacement) to models (in this case decision tree) each model performs independently & gives the result. The final result is checked through the decision of all the models. In classification through voting & average in regression.

Combining different Decision Tree to make predictions is a Random Forest.

Random forest reduces overfitting.

Bagging & Random forest reduce overfitting as.

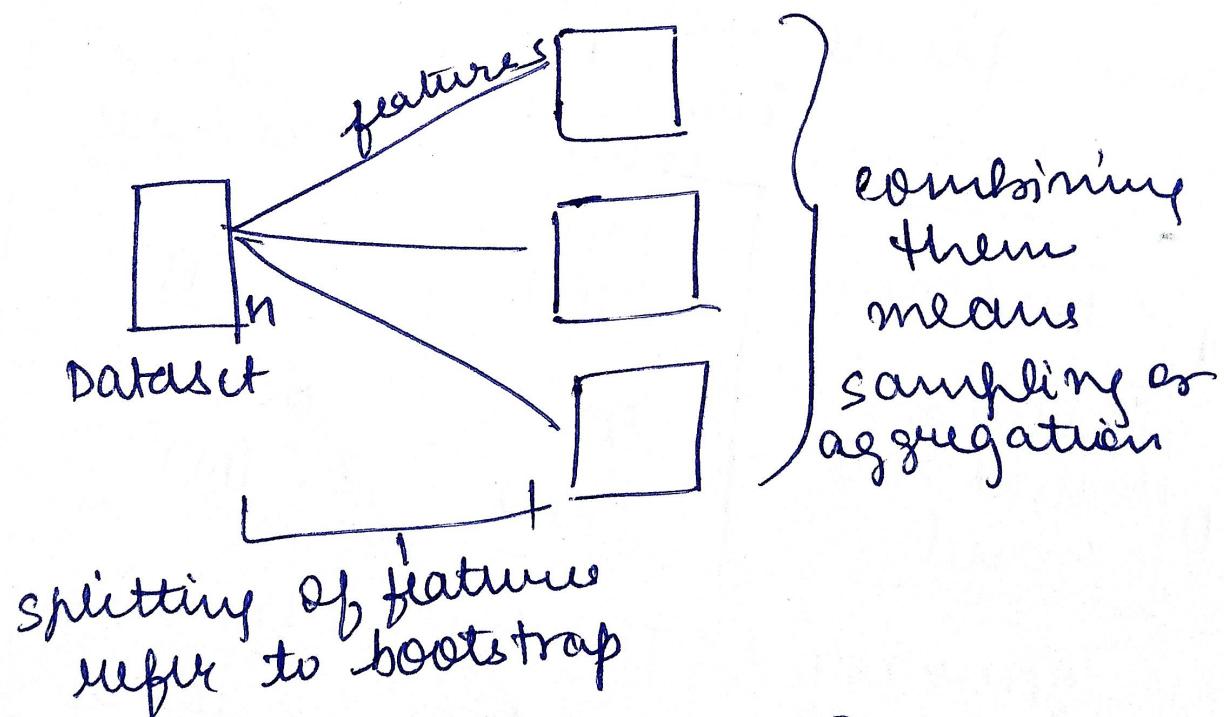
As features are distributed trees give different results thus repeating of feature problem is solved.

Q11 Decision trees have more depth. But if features are distributed many decision trees are built & hence reduce max depth of the depth & thus reduce overfitting.

Q12 Random Forest is an ensemble technique where various decision trees are used to make decisions & reduce overfitting.

→ Bootstrap Sampling

or simply Bagging means a dataset features are randomly selected & sent to different module for training & test with or without replacement.



→ Random feature selection

Features are randomly selected for each model.

without replacement - only new features are sent & features are not repeated