A.

How SVM works

The Line in the middle is called the hyperplane. The hyperplane has dotted lines on both sides called margins, which shouldn't be larger or smaller than necessary. The margins separate the two classes, instances on the margin are called support vectors. Slack variables are variables that allow some observations to fall on the wrong side of the margin to get better results. The C Hyperparameter controls how much slack is allowed to have on the determination of the optimal w. The larger the value of C, the larger the margin and more support vectors. The smaller the value of C the lower the bias and higher the variance.

How SVM kernels work

Linear kernel: only has constant C as a parameter and is usually used on data sets that have a lot of features. Used when data is linearly separable

Polynomial kernel: A linear classifier may not be able to classify data all the time so we can map the observations onto a higher dimensional space. A polynomial kernel can then be used to separate the observations.

Radial kernel: maps nonlinear data into a higher dimensional space similar to polynomial kernel. It has an additional hyperparameter called gamma which is then tuned on the validation data. A larger value of gamma can over fit which leads to a model with high variance. A smaller gamma value can under fit leading to a model with high bias.

Strengths and weaknesses of SVM

Advantages of svm :

- Be used for classification or regression
- Versatile because can choose from radial, polynomial and linear

Weaknesses of svm :

- computationally complex
- have to do tuning on hyperparameters to get the best result

B.

How Random forest works

Random forest generates many trees and then aggregates the best model from those trees. It also randomly tries a subset of variables at each split which decorrelates the trees. The parameter importance tells the random Forest algorithm to consider predictor importance and the parameter mtry tells how many variables to sample at each split. Random forest uses a parallel approach which means that it uses many weak learners that are created at the same time and the results are then aggregated.

How the other 2 algorithms used work

fastadaboost , uses c++ code to run a 100 times faster than r based libraries. Uses Freund and Schapire’s adaboost and also zhu’s samme algorithm.

XGBoost, the computation part is written in c++and can perform multithreading on one machine. Its first argument is the training matrix which doesn't include the label,cthe 2nd argument is the label, and the 3rd argument is the number of rounds.

Compared to these two algorithms, in a simple decision tree, input observations are recursively split into partitions till the observations in a given partition are uniform.

Strengths and weaknesses

Random Forest

Strengths of Random Forest

- Helps counteract the high variance in decision trees
- Overcomes the strong predictor

Weaknesses of Random Forest

- Takes a long time to run
- Looses interpretability

XGBoost

Strengths of XGBoost:

- has extreme scalability, running 10 times faster than earlier algorithms.
- Internally handles missing values
- Performs both regression and classification

Weakness of XGBoost:

- Training data has to be converted into a numeric matrix and Training labels have to be converted into 0 and 1

fastAdaboost

Strengths of fastAdaboost:

- Runs very quickly

Weakness of fastAdaboost:

- only works for binary classification problems.

.