Lab4 Report

Task 1:

Random Forest:

Experiments:

n_estimators	accuracy
5	77.40%
10	79.40%
20	78.07%
30	77.07%

Best model:

n_estimators	accuracy	Confusion matrix
10	79.40%	[221, 17], [45, 18]

AdaBoost:

Experiments:

n_estimators	accuracy
5	80.39%
10	82.39%
20	81.73%
30	81.40%

Best model:

n_estimators	accuracy	Confusion matrix
10	82.39%	[230, 8], [45, 18]

Results:

Random forest is an ensemble model using bagging as the ensemble method and decision tree as the individual model. AdaBoost is a boosting ensemble model and works especially well with the decision tree. Boosting model's key is learning from the previous mistakes, so AdaBoost has the better performance on this binary classification problem.

Task 2:

2.1:

Neural Network:

Best model:

Hidden layer	nodes	Learning rate	accuracy	Confusion matrix
1	4	0.1	81.40	[237 1] [55 8]

KNN algorithm:

Best model:

n_nerghbors	accuracy	Confusion matrix

10	81.73%	[230 8] [47 16]

Logistic regression:

Best model:

solver	Multi_class	accuracy	Confusion matrix
lbfgs	multinomial	80.73%	[232 6] [52 11]

Naïve Bayes:

Best model:

Likelihood model	accuracy	Confusion matrix
Gaussian	80.06%	[226 12] [48 15]

Decision Tree:

Best model:

accuracy	Confusion matrix
71.10%	[195 43] [44 19]

2.2:

Unweighted majority vote:

accuracy	Confusion matrix
82.39%	[230 8] [45 18]

2.3:

Weighted majority vote:

Retio(NN:KNN:LR:NB:DT)	accuracy	Confusion matrix
81 : 82: 81 : 80 : 71	83.39%	[230 8] [42 21]

Results:

The unweighted majority vote ensemble classifier has the similar performance with the best model (n_nergnbors) in the five models. The weighted majority vote ensemble classifier has the better performance than all the five models.

Task 3:

Unweighted majority vote:

accuracy	Confusion matrix
82.05%	[230 8] [48 15]

Weighted majority vote:

Retio(NN:KNN:LR:NB:DT:RF:ADA)	accuracy	Confusion matrix
81 : 82 : 81 : 80 : 71 : 79 : 82	84.05%	[229 9] [39 24]

Results:

The unweighted majority vote ensemble classifier has the similar performance with the best model (AdaBoost) in the seven models. The weighted majority vote ensemble classifier has the better performance than all the seven models and also better than the weighted majority vote ensemble model in task 2.