PS10

Donald Li

February. 2018

Worked with Jordan and Alex

1 Part 7

From tuning model

Tree
minsplit=10;
minbucket=6;
cp=0.026 :
f1.test.mean=0.889,
gmean.test.mean=0.661

Logit
lambda=0.0243;
alpha=0.746 :
f1.test.mean=0.90
,gmean.test.mean=0.7

NN
size=9;
decay=0.155;
maxit=1000 :
f1.test.mean=0.90
,gmean.test.mean=0.774

KNN
k=23 :
f1.test.mean=0.897
,gmean.test.mean=0.747

SVM
cost=1;
gamma=0.5

: f1.test.mean=0.90,
gmean.test.mean=0.69

2 Part 8 and 9

Verify the performance on cross validated sets:

Model names	f1	gmean	Out-of-sample f1	Out-of-sample gmean
Tree	0.896	0.658	0.8968421	0.6730932
Logit	0.897	0.662	0.8986422	0.6762722
NN	0.906	0.765	0.9094500	0.7675114
KNN	0.901	0.751	0.8975970	0.7564945
SVM	0.906	0.732	0.9048742	0.7478175
NB	0.884	0.726	0.8825952	0.7340489

All the models preformed similarly. The two that stand out are the Tree and Logit models. I would chose the Tree model over Logit but that is an arbitrary decision, either models will do.