

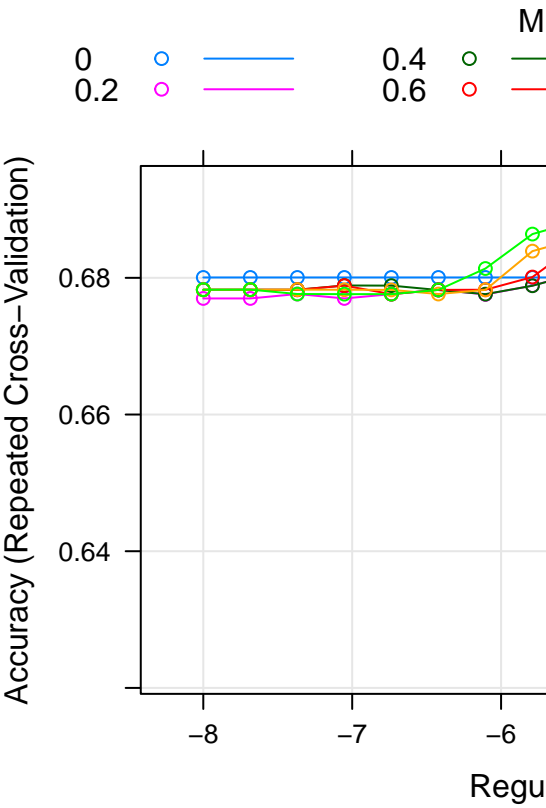
models

Elaine Xu

Contents

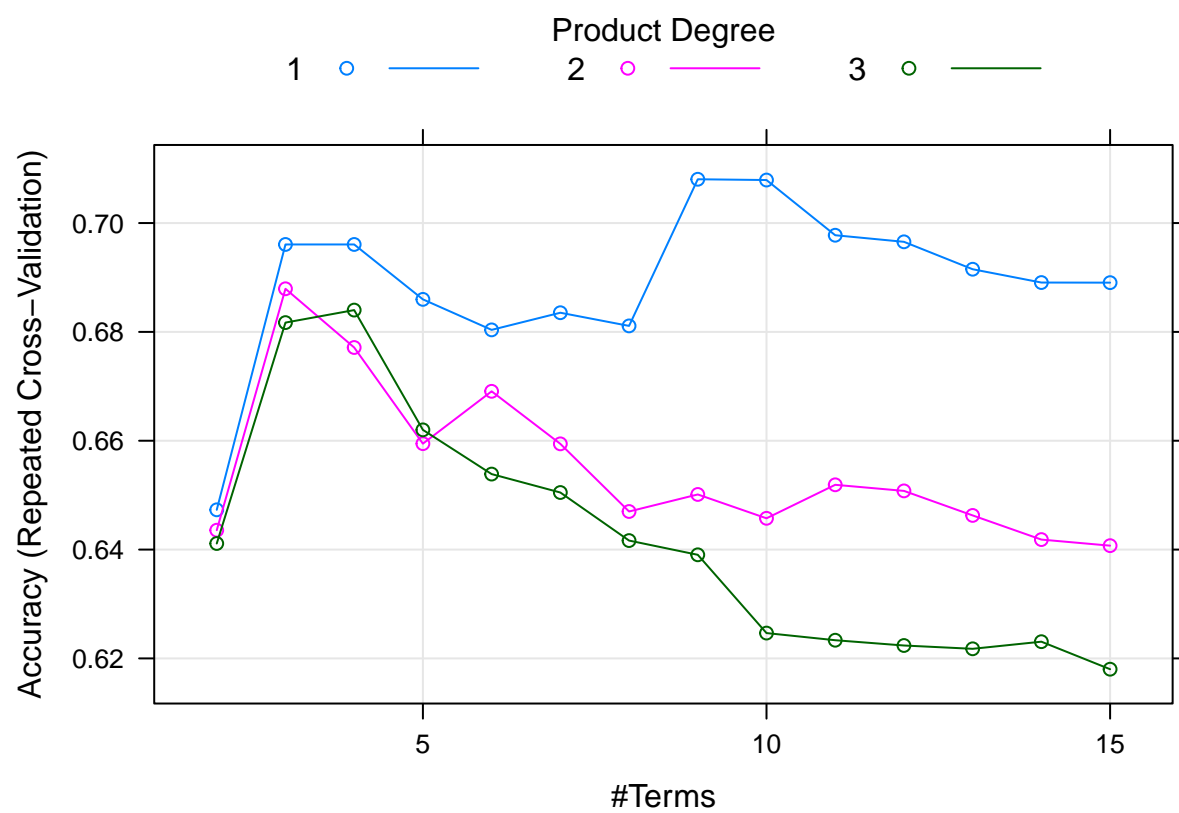
Models	1
Model comparison	9
Result	10

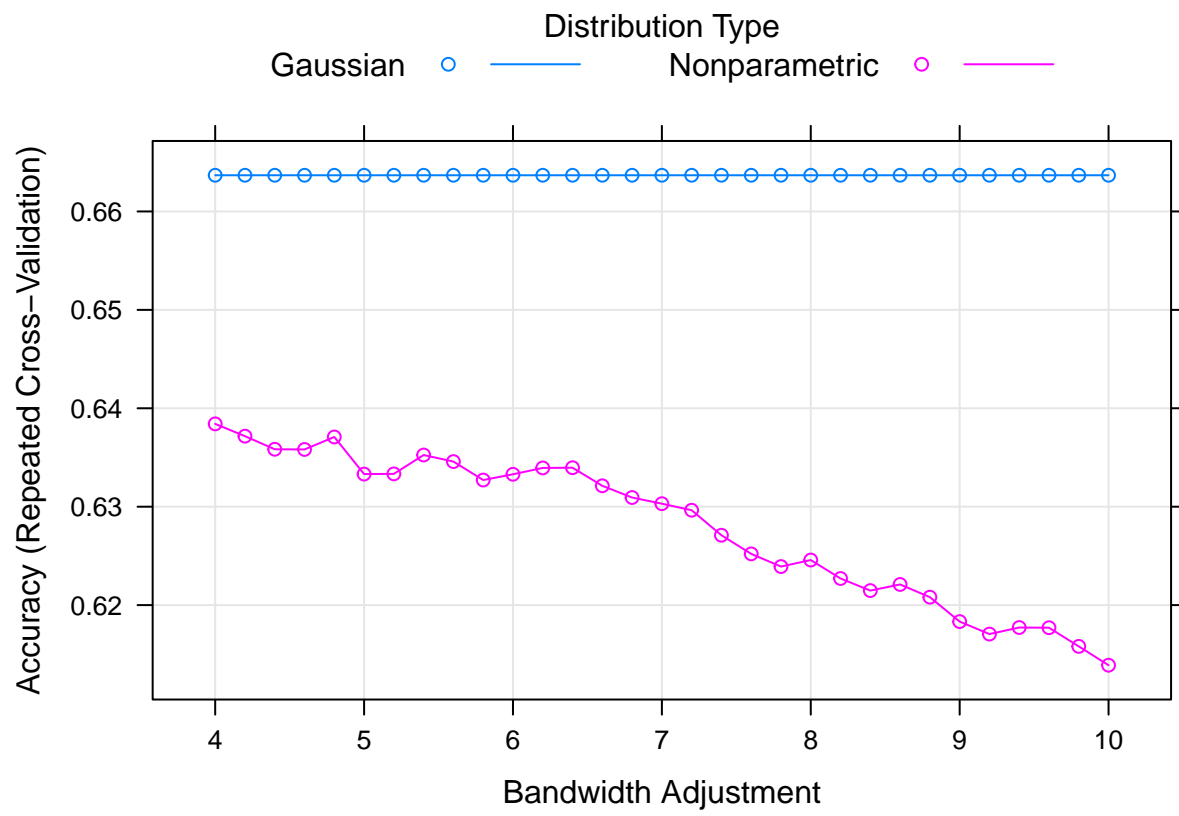
Models



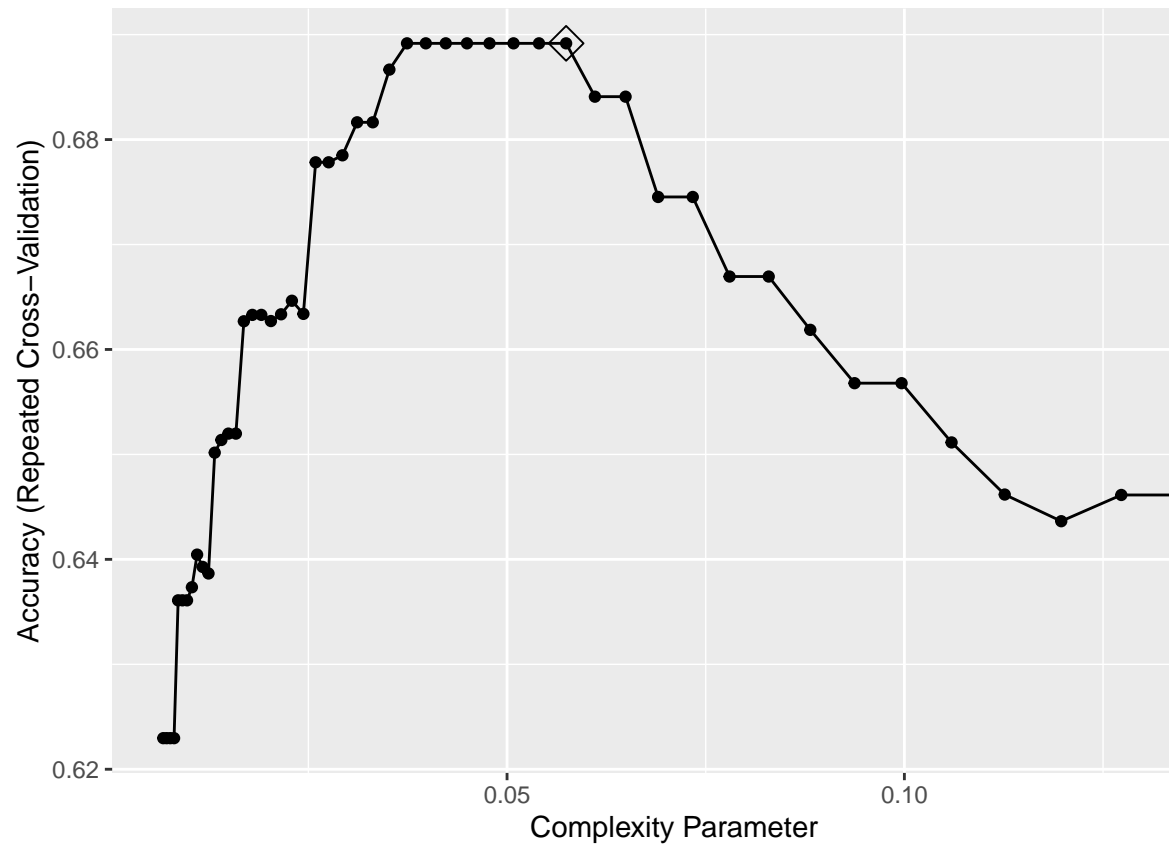
Linear methods: glm, penalized logistic regression, GAM, MARS

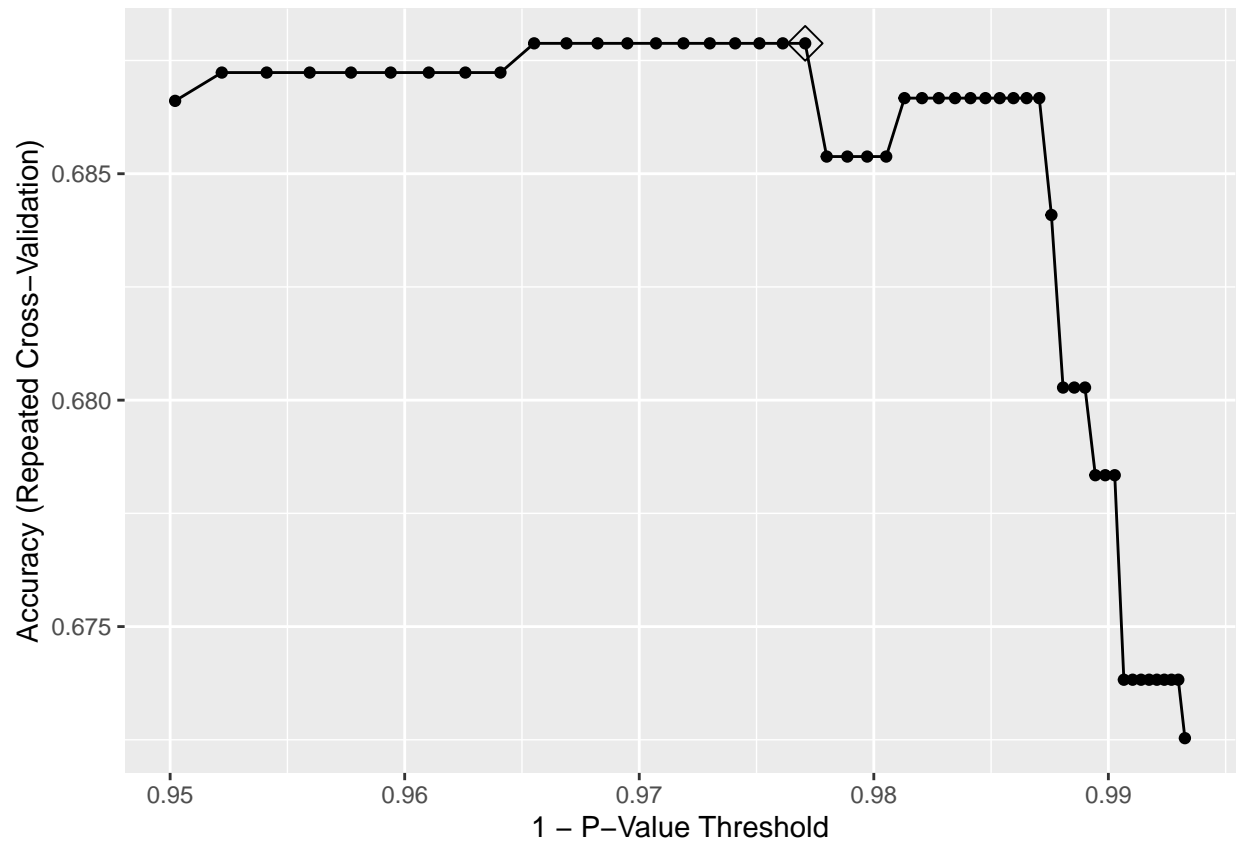
```
##      alpha      lambda
## 59    0.4 0.09868824
```



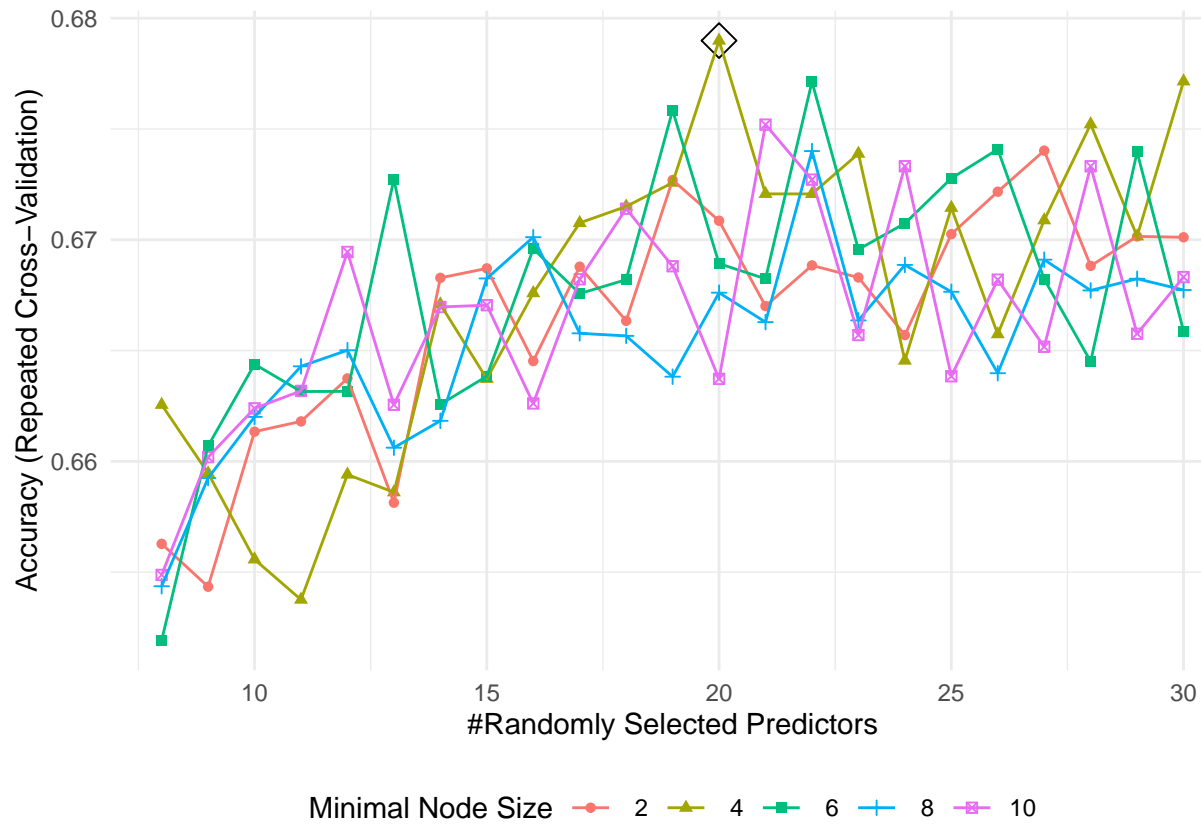


LDA/QDA/NB



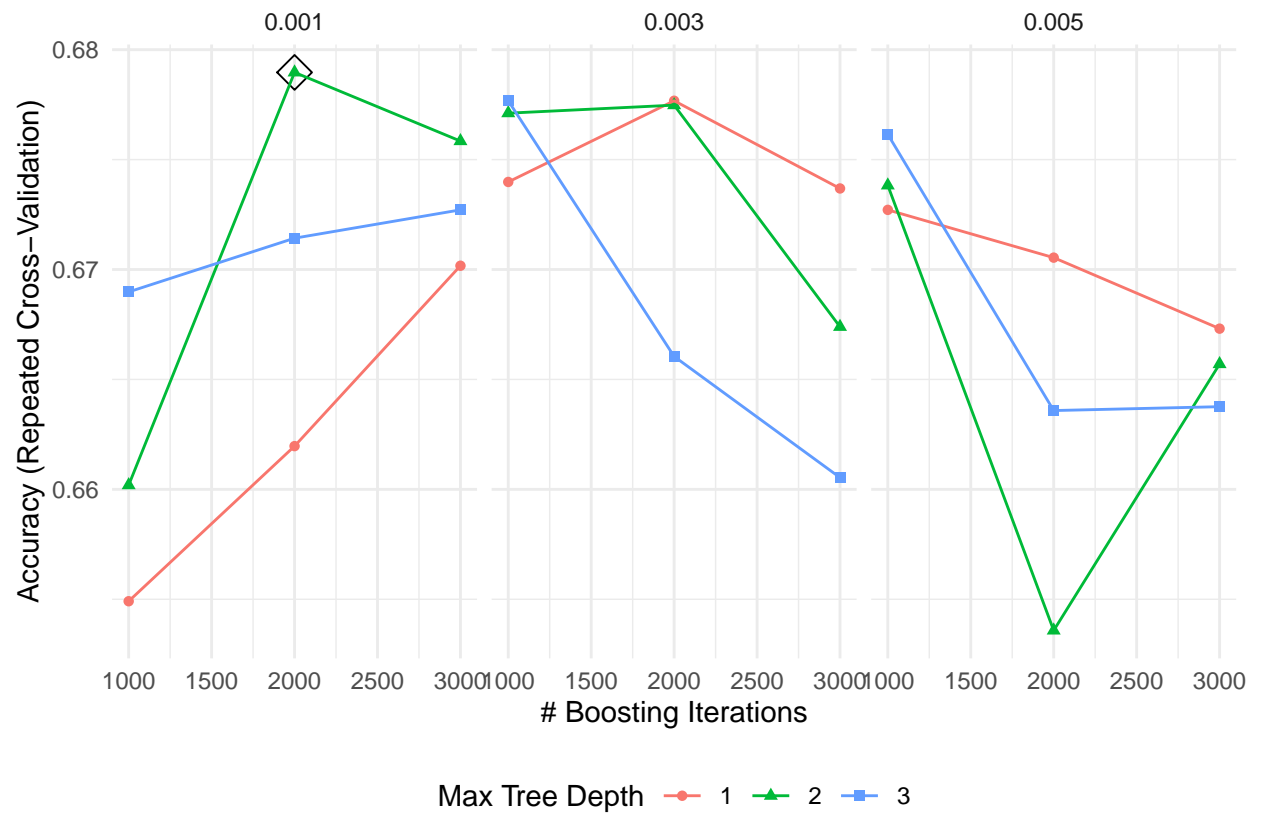


```
## mincriterion
## 20 0.9770746
```

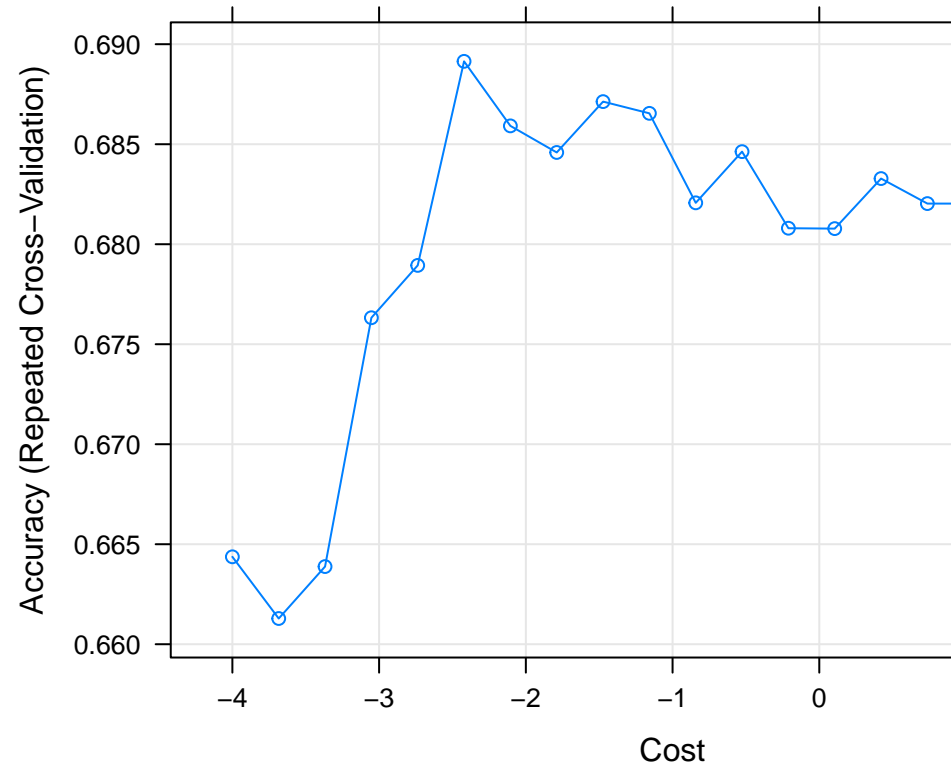


Random Forest

```
##      mtry splitrule min.node.size
## 62    20      gini              4
```

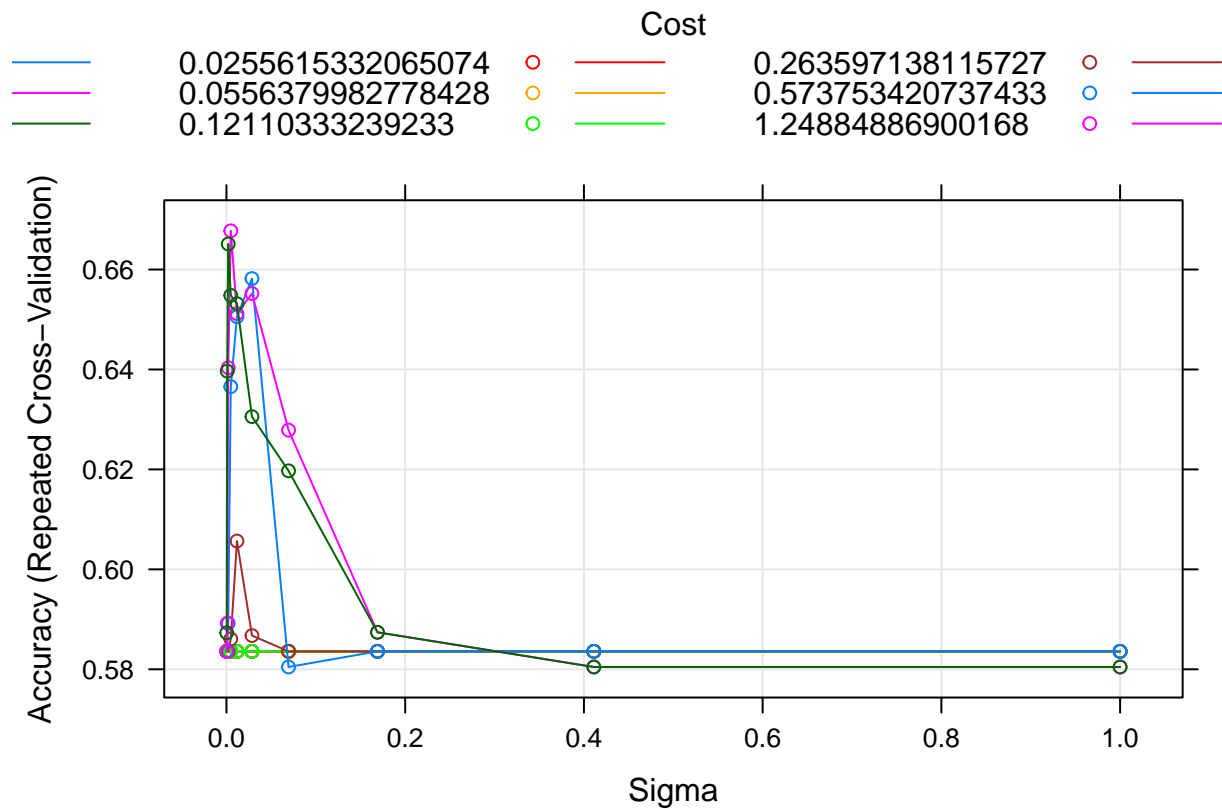


Adaboost



Support vector classifier and SVM

```
## C
## 6 0.08882807
```

```
##      sigma      C
## 84 0.00482795 1.248849
```

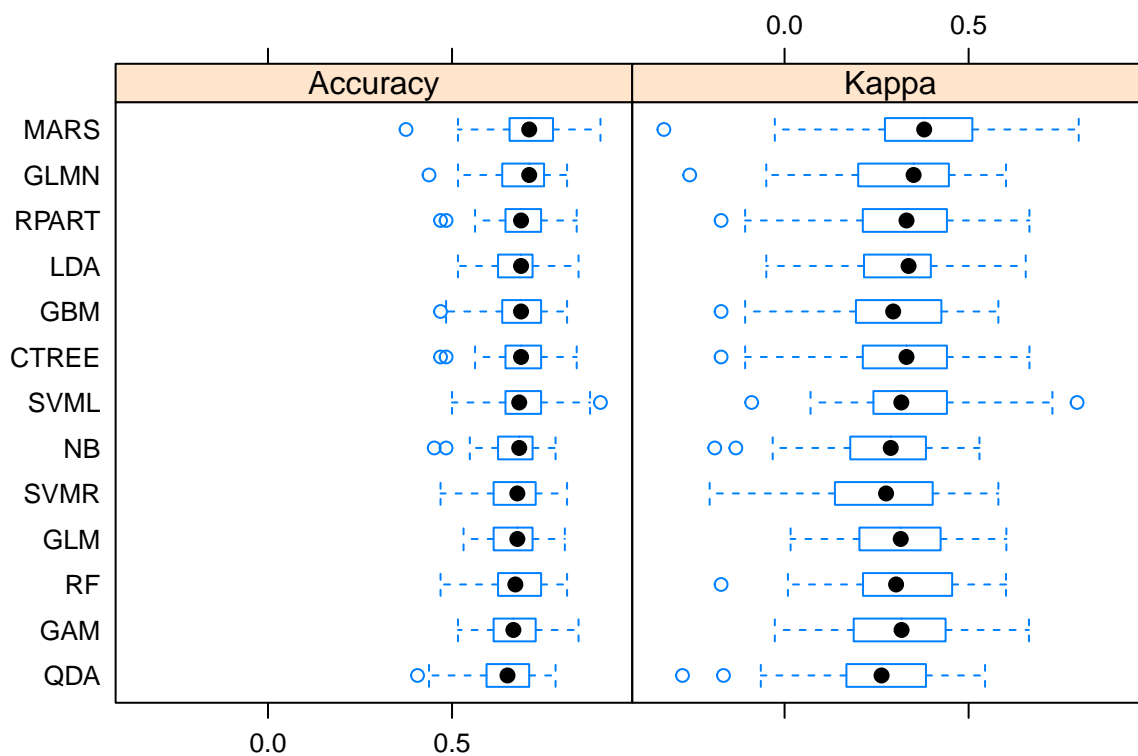
Model comparison

```
##
## Call:
## summary.resamples(object = res)
##
## Models: GLM, GLMN, GAM, MARS, LDA, QDA, NB, RPART, CTREE, RF, GBM, SVML, SVMR
## Number of resamples: 50
##
## Accuracy
##      Min.   1st Qu.   Median     Mean   3rd Qu.     Max. NA's
## GLM    0.5312500 0.6129032 0.6774194 0.6731940 0.7187500 0.8064516    0
## GLMN   0.4375000 0.6385630 0.7096774 0.6916288 0.7479839 0.8125000    0
## GAM    0.5161290 0.6129032 0.6666667 0.6725122 0.7251420 0.8437500    0
## MARS   0.3750000 0.6588542 0.7096774 0.7080486 0.7681452 0.9032258    0
## LDA    0.5161290 0.6250000 0.6875000 0.6745424 0.7187500 0.8437500    0
## QDA    0.4062500 0.5937500 0.6507056 0.6408242 0.7096774 0.7812500    0
## NB     0.4516129 0.6250000 0.6824597 0.6636779 0.7164819 0.7812500    0
## RPART  0.4687500 0.6479335 0.6875000 0.6891691 0.7419355 0.8387097    0
## CTREE  0.4687500 0.6479335 0.6875000 0.6878788 0.7419355 0.8387097    0
## RF     0.4687500 0.6300403 0.6720430 0.6789889 0.7382698 0.8125000    0
## GBM    0.4687500 0.6385630 0.6875000 0.6789663 0.7382698 0.8125000    0
```

```

## SVML 0.5000000 0.6451613 0.6824597 0.6891428 0.7382698 0.9032258 0
## SVMR 0.4687500 0.6129032 0.6774194 0.6677517 0.7251420 0.8125000 0
##
## Kappa
##      Min.    1st Qu.    Median      Mean    3rd Qu.      Max. NA's
## GLM      0.01639344 0.2064202 0.3155899 0.3139664 0.4204590 0.6025641 0
## GLMN     -0.25764192 0.2000000 0.3505564 0.3162002 0.4464286 0.6016598 0
## GAM      -0.02649007 0.1877729 0.3178221 0.3168699 0.4372332 0.6638655 0
## MARS     -0.32780083 0.2785455 0.3793103 0.3773783 0.4998384 0.7991361 0
## LDA      -0.04966140 0.2174596 0.3368532 0.3096634 0.3967957 0.6551724 0
## QDA      -0.27731092 0.1712756 0.2634989 0.2465061 0.3829071 0.5447154 0
## NB       -0.18961625 0.1795621 0.2883951 0.2676388 0.3829071 0.5294118 0
## RPART    -0.17241379 0.2146631 0.3313008 0.3275515 0.4392335 0.6652268 0
## CTREE    -0.17241379 0.2146631 0.3313008 0.3246706 0.4392335 0.6652268 0
## RF       -0.17241379 0.2131148 0.3029540 0.3140537 0.4532212 0.6016598 0
## GBM      -0.17241379 0.1953901 0.2950549 0.2922473 0.4258897 0.5807860 0
## SVML     -0.08936170 0.2428446 0.3171698 0.3404805 0.4371690 0.7947020 0
## SVMR     -0.20353982 0.1402869 0.2757009 0.2566076 0.3943617 0.5807860 0

```



Result

```

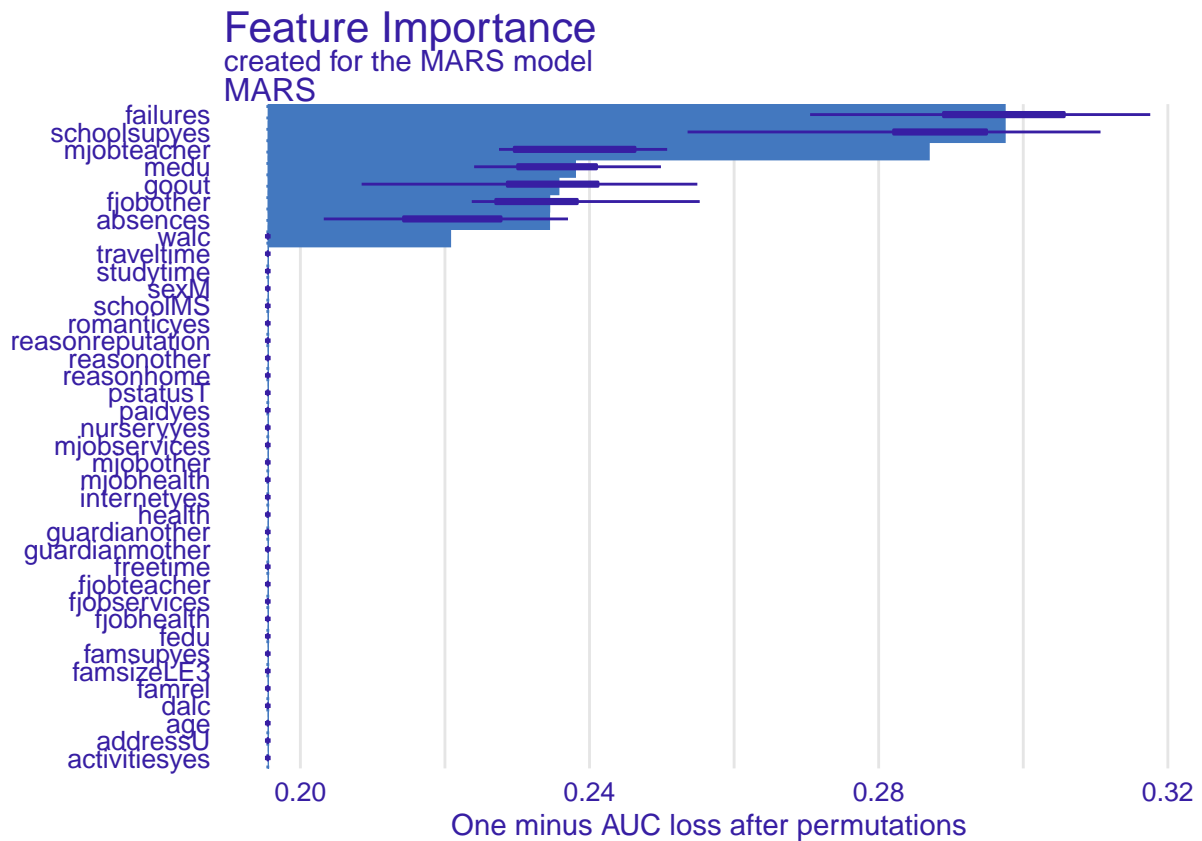
## (Intercept) h(failures-1) h(1-failures) schoolsupyes h(4-goout)
## 0.7133150 -2.0783188 1.2018131 -2.1353865 0.5488283
## h(2-absences) h(3-medu) mjobteacher fjobother
## -0.4227162 -0.6986678 -1.4430923 -0.9516230

```

Since MARS was the best model with the best prediction accuracy demonstrated by the resampling comparison, we examined its test data set performance. We obtained a test accuracy of 62.82%.

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction fail pass
##      fail   13   10
##      pass   19   36
##
##           Accuracy : 0.6282
##           95% CI : (0.5113, 0.735)
##      No Information Rate : 0.5897
##      P-Value [Acc > NIR] : 0.2842
##
##           Kappa : 0.1973
##
##  McNemar's Test P-Value : 0.1374
##
##           Sensitivity : 0.4062
##           Specificity : 0.7826
##           Pos Pred Value : 0.5652
##           Neg Pred Value : 0.6545
##           Prevalence : 0.4103
##           Detection Rate : 0.1667
##      Detection Prevalence : 0.2949
##           Balanced Accuracy : 0.5944
##
##           'Positive' Class : fail
##
```

We further examined variable importance, and we can see that “failures”, “schoolsupyes” and “goout” were the top three important features in making the prediction.

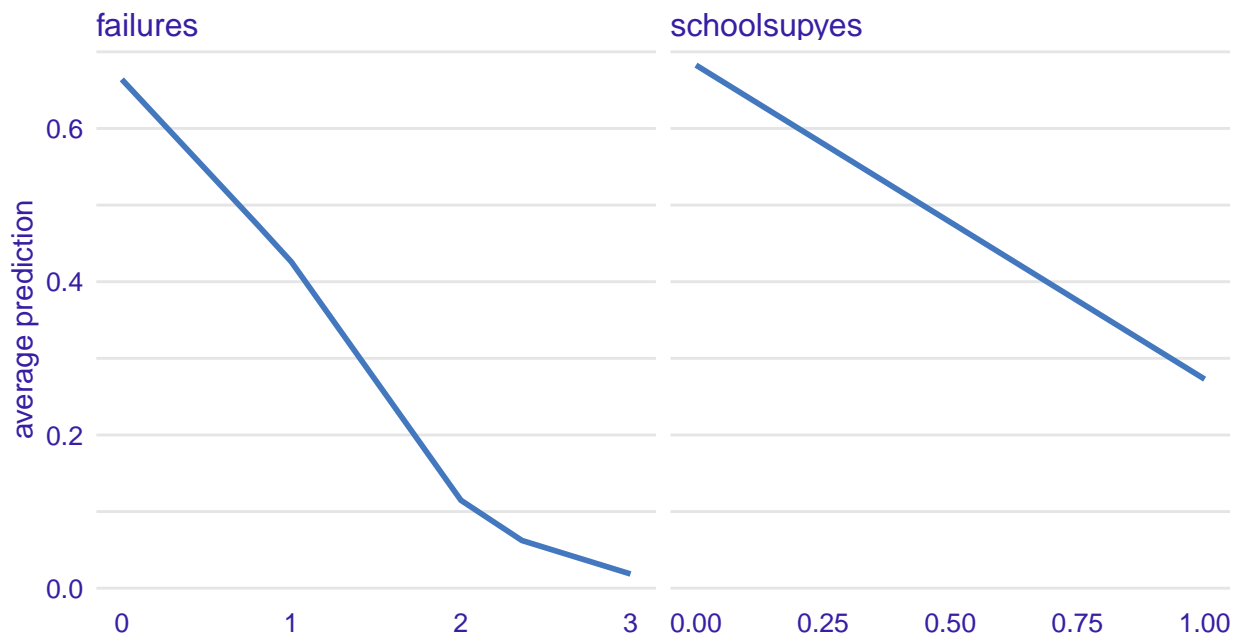


We also made partial dependence profile for “failures” and “schoolsupyes”. As demonstrated by the plots, the average probability of “pass” decreases as failures increases and for students without extra educational school

Partial Dependence profile

Created for the MARS model

— MARS



support.