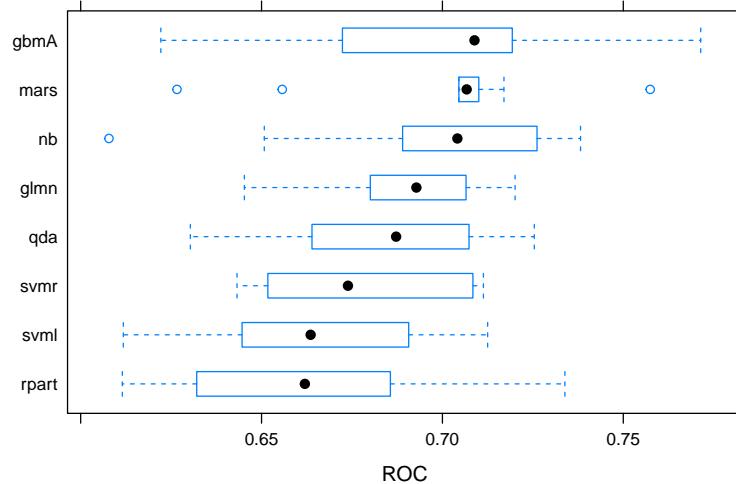


Figures and Tables for Secondary Analysis

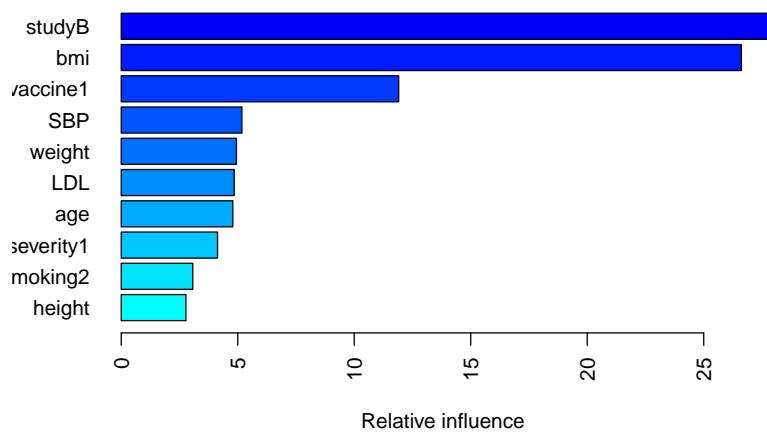
Models Comparsion based on Area Under Curves



Final model performance on test dataset

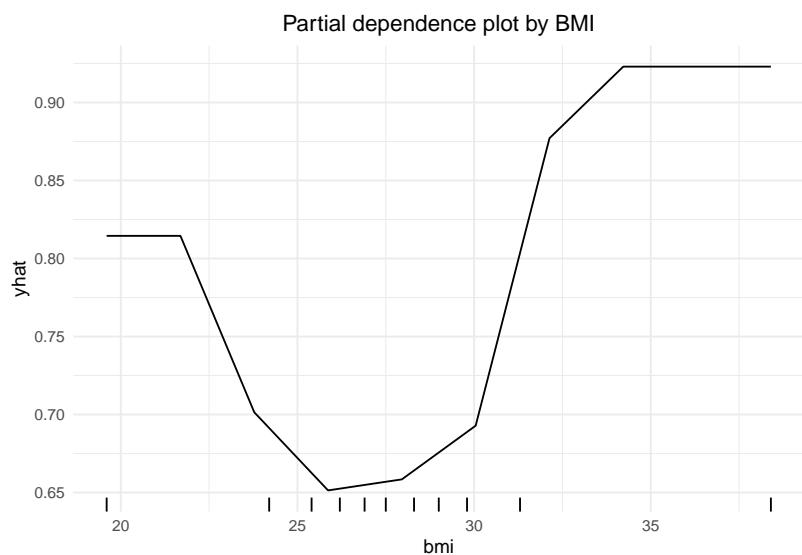
| x | |
|----------|-----------|
| Accuracy | 0.7218359 |
| Kappa | 0.1815036 |

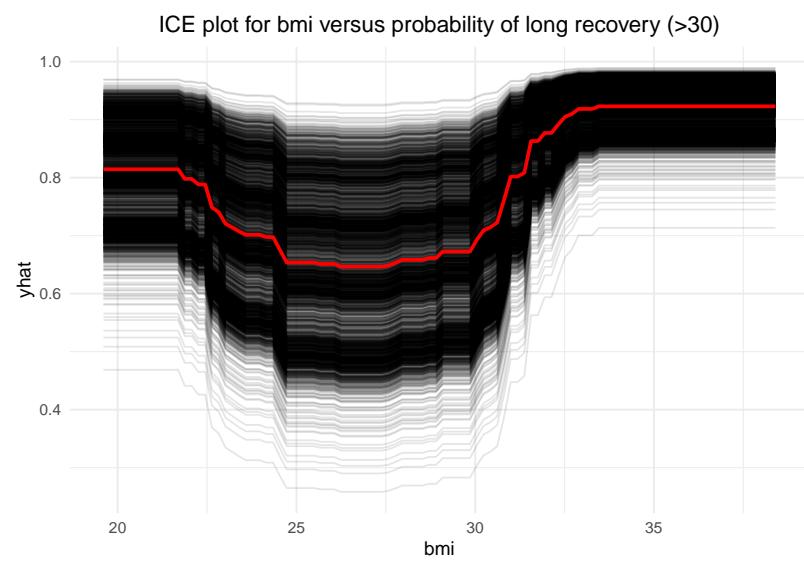
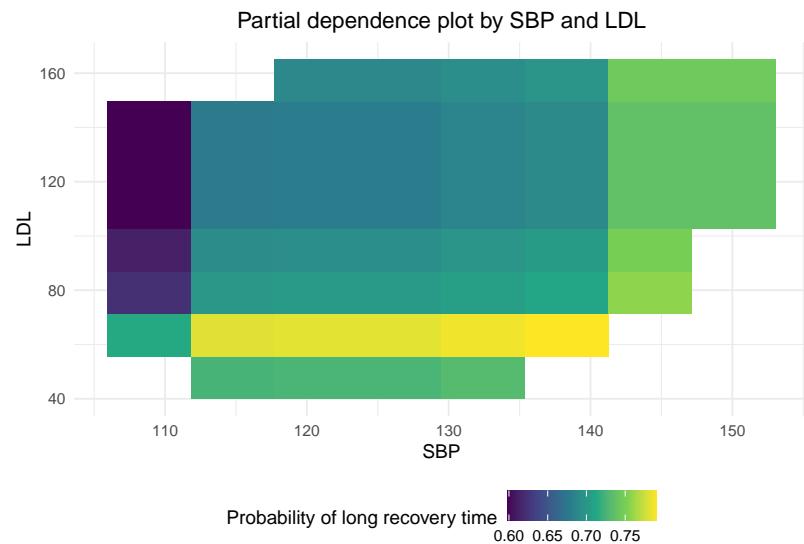
Variable importance plots of the final model



| | rel.inf |
|---------------|------------|
| studyB | 27.8958031 |
| bmi | 26.6131604 |
| vaccine1 | 11.9080853 |
| SBP | 5.1783551 |
| weight | 4.9451667 |
| LDL | 4.8495707 |
| age | 4.7895698 |
| severity1 | 4.1331856 |
| smoking2 | 3.0702806 |
| height | 2.7755020 |
| gender1 | 1.1978980 |
| smoking1 | 1.0717996 |
| race4 | 1.0470291 |
| race2 | 0.2303987 |
| race3 | 0.1567322 |
| studyC | 0.1374633 |
| hypertension1 | 0.0000000 |
| diabetes1 | 0.0000000 |

Final Model interpretation

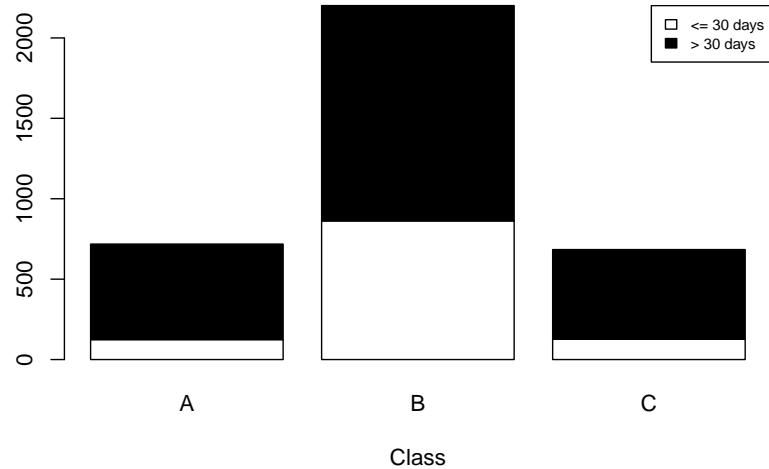




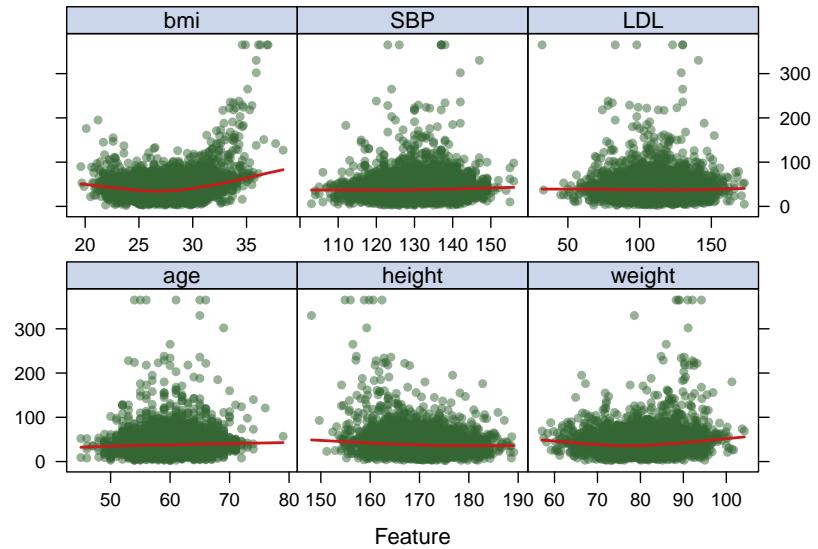
Appendix – Primary Analysis of Covid Recovery Time

Exploratory analysis and data visualization

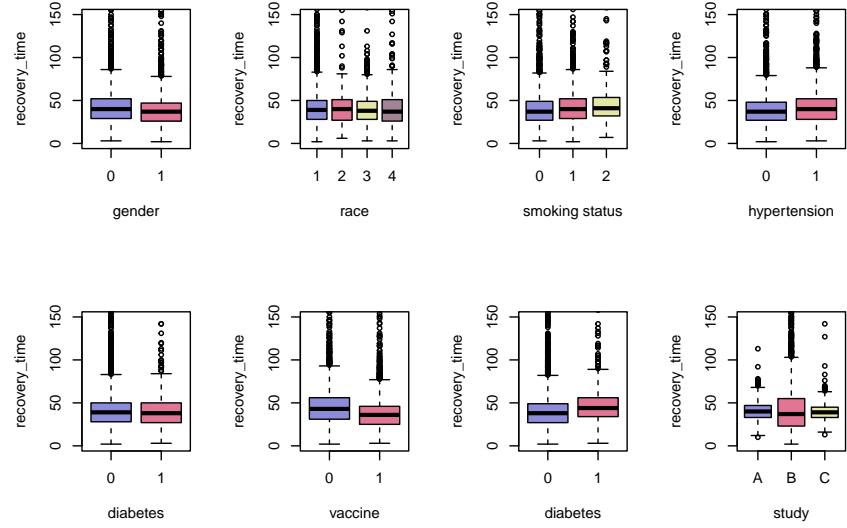
Number of cases seperated by 30 days in recovery by Study Group:



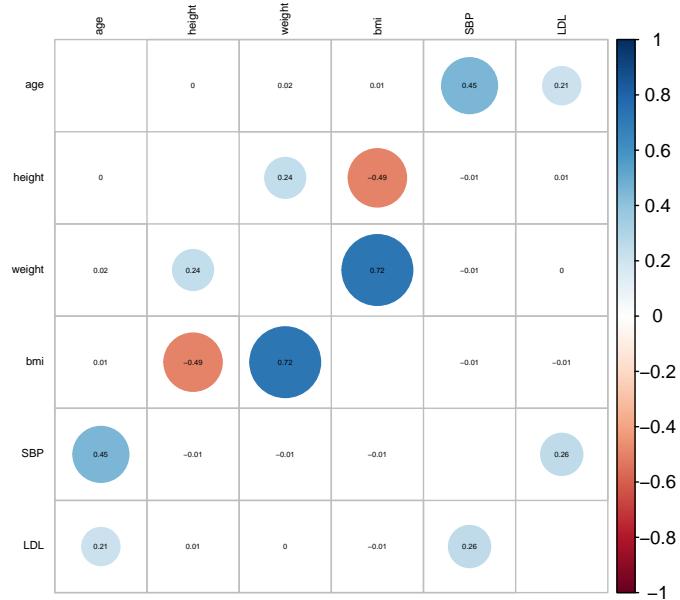
Visualize potential relationship between reponse variable and numeric predictors



Visualize potential relationship between response variable and categorical predictors

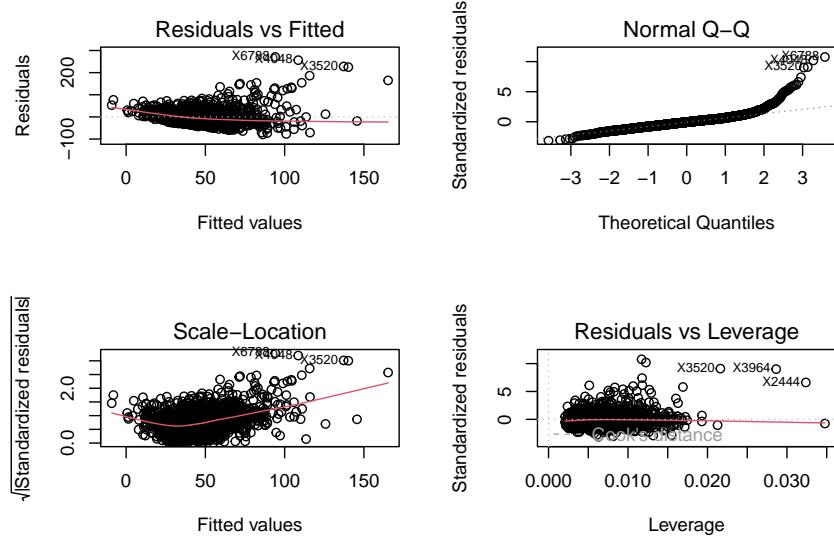


Correlation plot to check collinearity between covariates (based on training data)

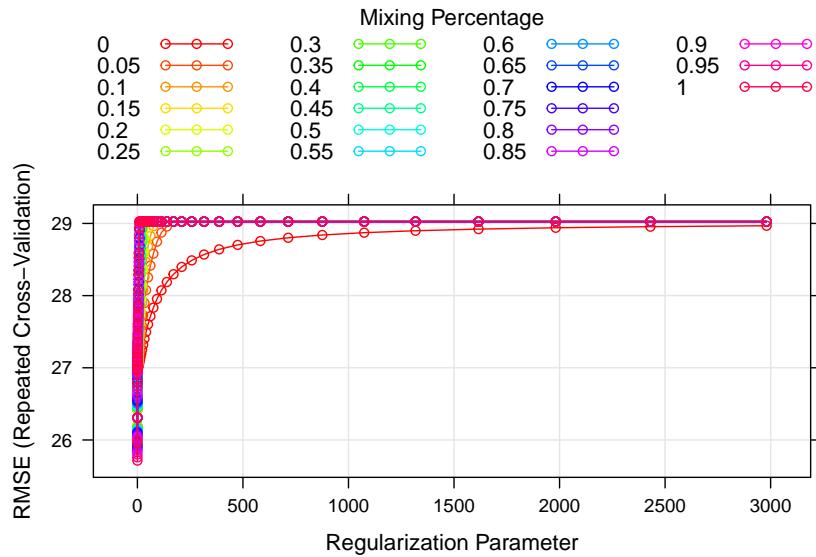


Model Training

Ordinary Least square

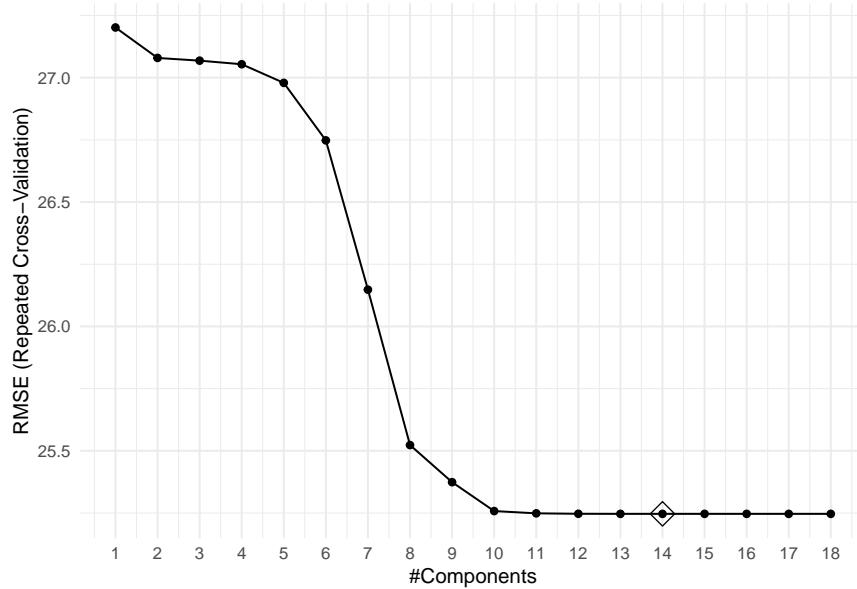


Elastic net regression

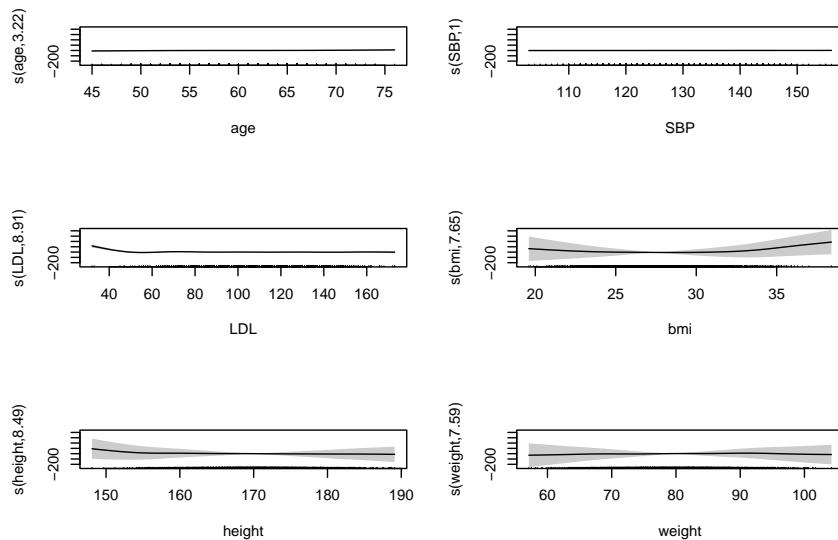


```
##      alpha     lambda
## 1001     1 0.1353353
```

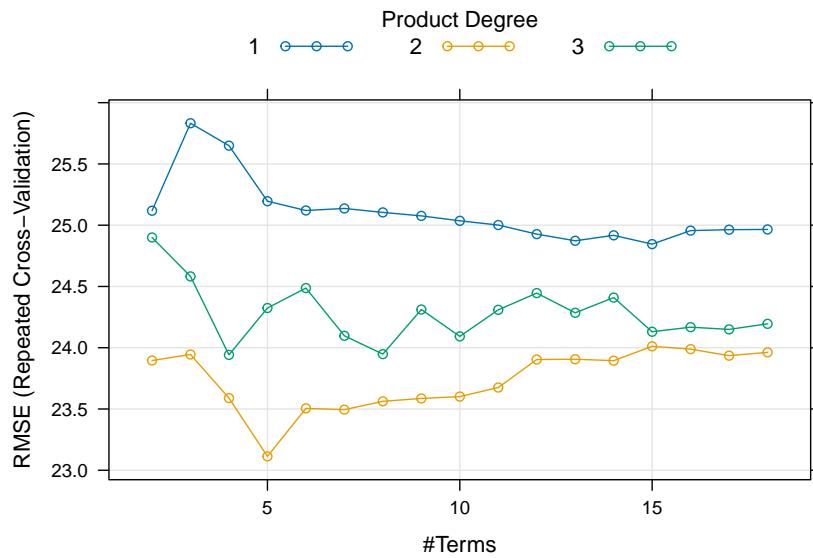
Partial least squares (PLS)



Generalized Additive Models (GAM)



Multivariate adaptive regression spline model (MARS)

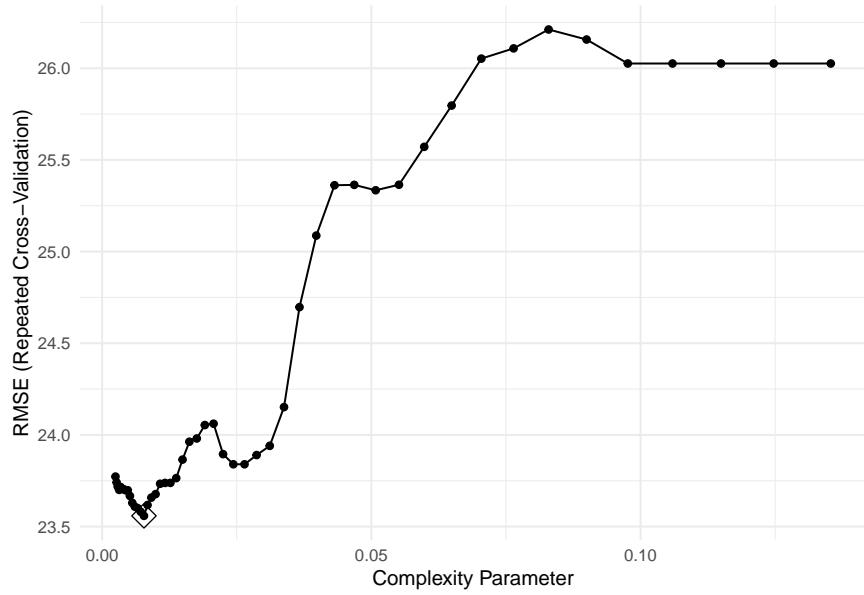


```
## Call: earth(x=matrix[2884,18], y=c(17,33,92,20,4...), keepxy=TRUE, degree=2,
##             nprune=5)
##
##                               coefficients
## (Intercept)              -2.941972
## vaccine1                 -9.506054
## h(bmi-24.5)                7.647737
## h(30.6-bmi)                7.320881
## h(bmi-30.6) * studyB    20.811559
##
## Selected 5 of 24 terms, and 3 of 18 predictors (nprune=5)
## Termination condition: Reached nk 37
## Importance: bmi, studyB, vaccine1, age-unused, gender1-unused, ...
## Number of terms at each degree of interaction: 1 3 1
## GCV 516.0731      RSS 1477023      GRSq 0.4041199      RSq 0.4082465
```

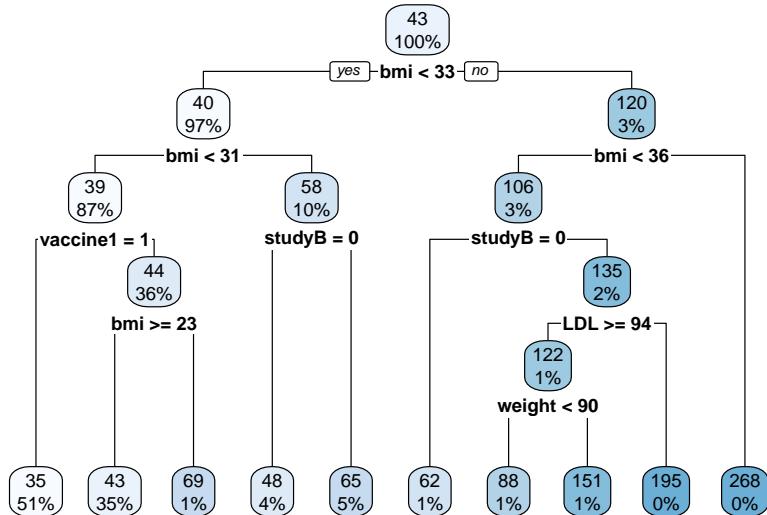
K-Nearest Neighbors (KNN)

```
##     k
## 9 13
```

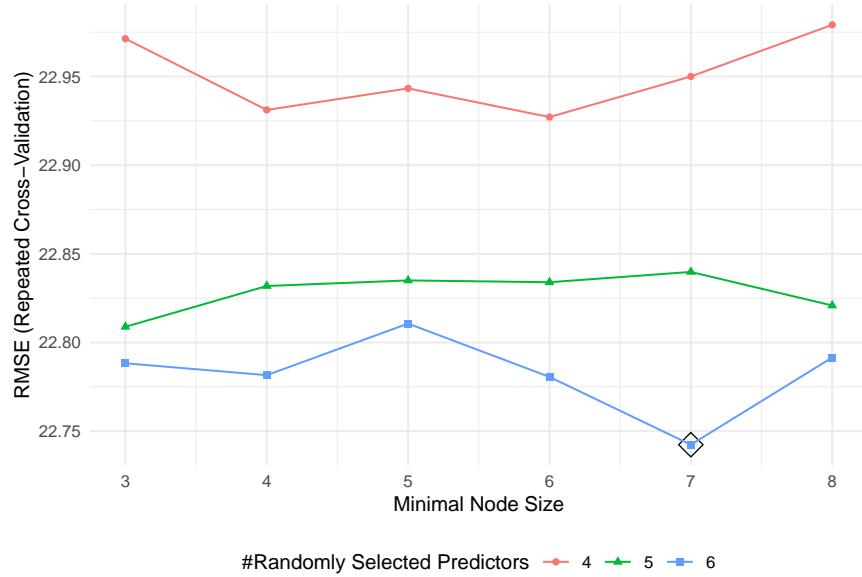
Regression Tree (CART)



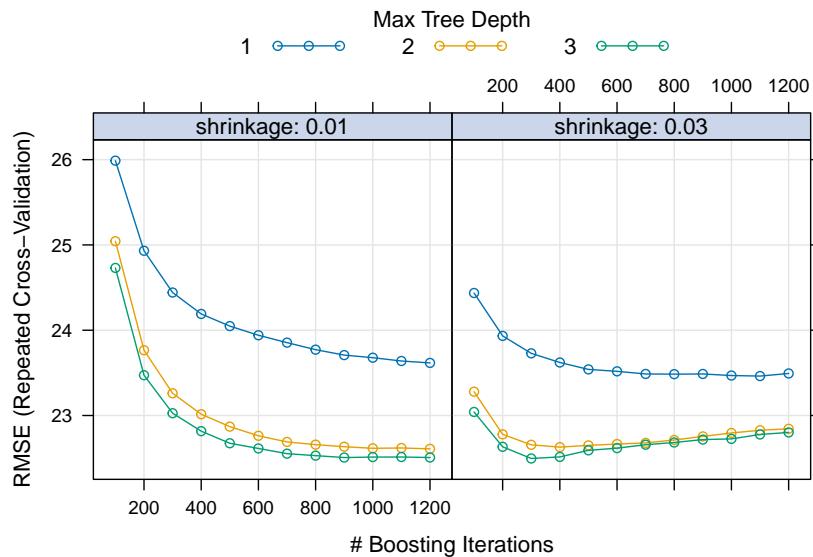
```
##          cp
## 15 0.00777266
```



Random Forest



Generalized Boosted Regression



Models comparsion based on cross validation error

```
## 
## Call:
## summary.resamples(object = resamp)
## 
## Models: lm, enet, pls, gam, mars, knn, gbm, cart, rf
## Number of resamples: 50
```

```

##  

## MAE  

##          Min. 1st Qu. Median   Mean 3rd Qu.   Max. NA's  

## lm    15.16547 16.10302 16.78177 16.83446 17.42209 19.04602 0  

## enet  14.56452 15.92803 16.52216 16.46752 17.01105 18.64227 0  

## pls   15.16548 16.10302 16.78177 16.83446 17.42208 19.04604 0  

## gam   14.35229 15.33438 15.83680 15.95120 16.56241 17.79521 0  

## mars  13.93931 14.88663 15.26803 15.37548 15.95721 17.55273 0  

## knn   13.77991 14.91692 15.66110 15.73718 16.40355 18.31290 0  

## gbm   13.16340 14.34295 14.85952 14.96472 15.45562 16.75934 0  

## cart  14.24679 14.87635 15.29857 15.47210 16.11928 17.75836 0  

## rf    13.66061 14.41903 14.83788 15.08850 15.72662 17.32203 0  

##  

## RMSE  

##          Min. 1st Qu. Median   Mean 3rd Qu.   Max. NA's  

## lm    19.74488 23.43892 25.17216 25.24668 27.21931 32.18024 0  

## enet  19.46546 23.14204 25.30888 25.71390 27.81579 33.96273 0  

## pls   19.74490 23.43892 25.17216 25.24668 27.21931 32.18025 0  

## gam   19.14744 22.27447 23.50243 24.21183 26.16733 31.05932 0  

## mars  18.46040 21.55759 22.41118 23.11308 24.68399 27.78102 0  

## knn   18.37738 23.09791 24.94296 26.01258 29.24863 36.95880 0  

## gbm   18.45737 21.10448 21.86910 22.49591 24.18224 27.22898 0  

## cart  18.92266 21.53195 22.80303 23.55843 25.57517 30.80530 0  

## rf    17.98350 21.08297 22.32057 22.74234 24.74996 28.19421 0  

##  

## Rsquared  

##          Min. 1st Qu. Median   Mean 3rd Qu.   Max. NA's  

## lm    0.08620897 0.1900059 0.2572871 0.2531044 0.3099187 0.4654627 0  

## enet  0.08078819 0.1811701 0.2277041 0.2290413 0.2731629 0.3775725 0  

## pls   0.08620900 0.1900061 0.2572879 0.2531045 0.3099185 0.4654625 0  

## gam   0.12269099 0.2416508 0.3294108 0.3350657 0.4183874 0.6223217 0  

## mars  0.07601256 0.2807337 0.3859944 0.3750641 0.4765129 0.6877643 0  

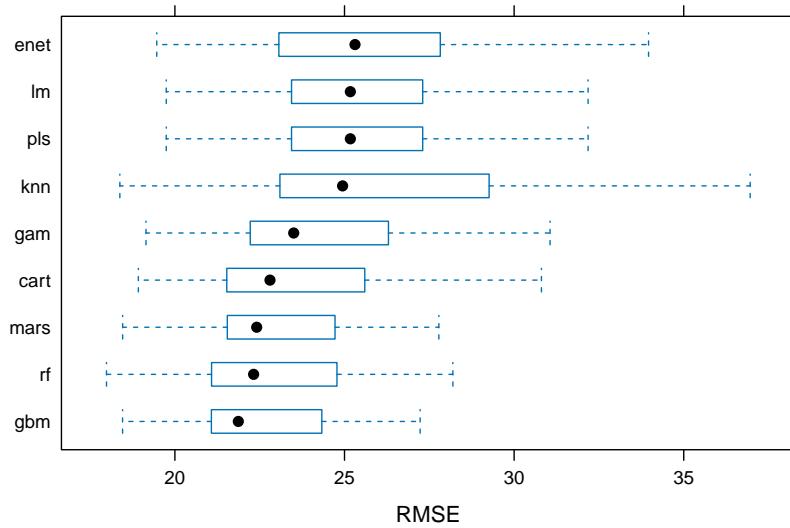
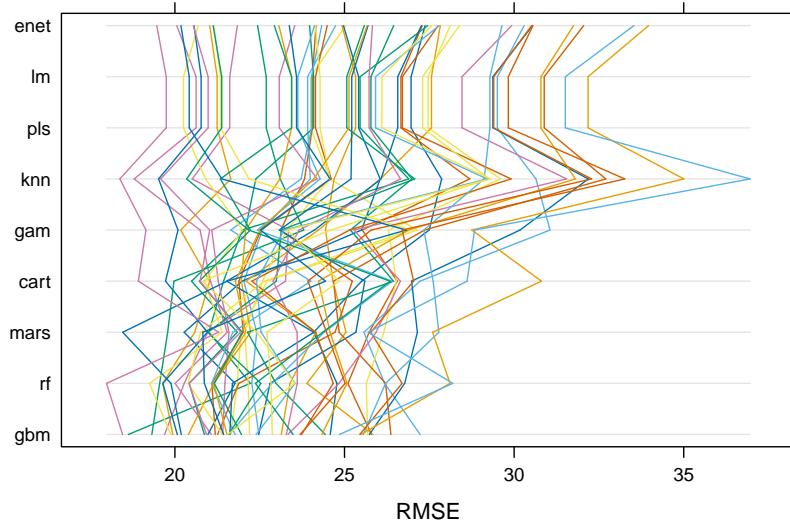
## knn   0.06268675 0.1685928 0.2095025 0.2280360 0.2800124 0.4727332 0  

## gbm   0.07050574 0.3111809 0.4021608 0.4026223 0.4940622 0.7041863 0  

## cart  0.05216633 0.2359045 0.3635301 0.3517752 0.4471816 0.6223333 0  

## rf    0.09295909 0.3246594 0.3947281 0.3929464 0.4732256 0.6063936 0

```

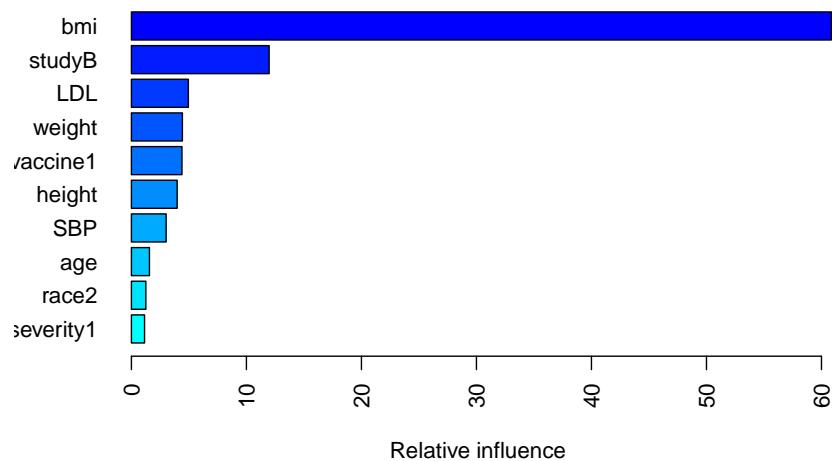


Results

Test Mean Squared Error

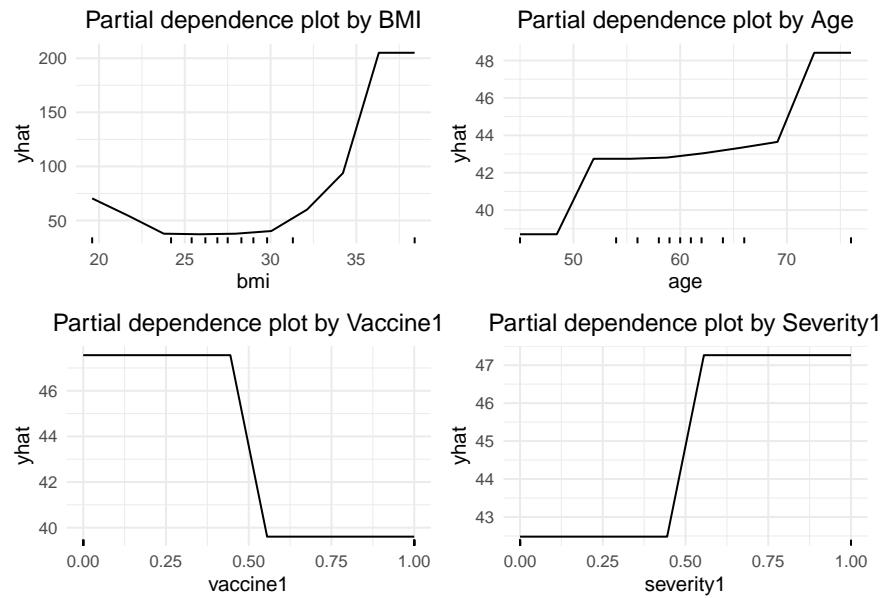
```
## [1] 492.5701
```

Variable importance plots

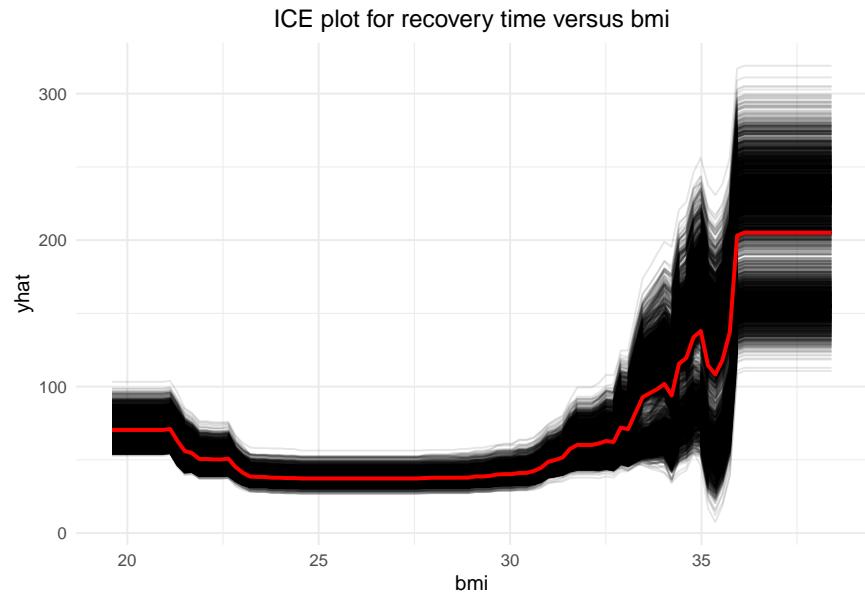


| | rel.inf |
|---------------|------------|
| bmi | 60.8581546 |
| studyB | 11.9739399 |
| LDL | 4.9517050 |
| weight | 4.4347751 |
| /vaccine1 | 4.3964715 |
| height | 3.9777372 |
| SBP | 3.0197676 |
| age | 1.5690459 |
| race2 | 1.2559401 |
| severity1 | 1.1426109 |
| gender1 | 1.0789580 |
| smoking2 | 0.5346932 |
| smoking1 | 0.3760449 |
| hypertension1 | 0.1605474 |
| diabetes1 | 0.1503396 |
| race4 | 0.1009740 |
| race3 | 0.0182951 |
| studyC | 0.0000000 |

Partial dependence plots



Individual Conditional Expectation (ICE) plot



Secondary Analysis – Figures and outputs

Exploratory analysis for classification

Table 1: Data summary

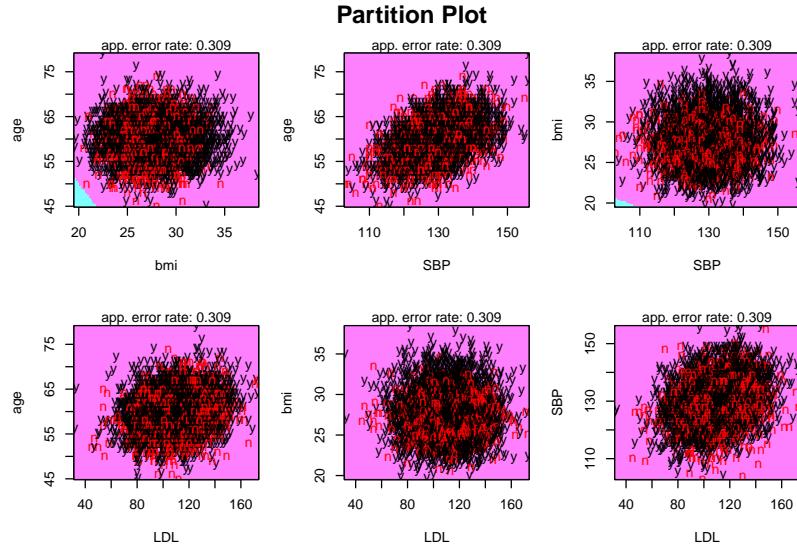
| | |
|------------------------|------------|
| Name | Piped data |
| Number of rows | 3603 |
| Number of columns | 17 |
| Column type frequency: | |
| factor | 9 |
| numeric | 8 |
| Group variables | None |

Variable type: factor

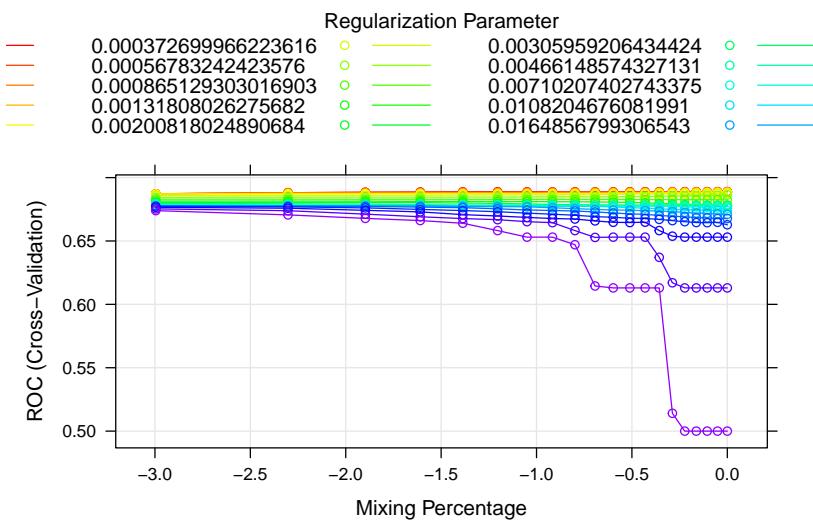
| skim_variable | n_missing | complete_rate | ordered | n_unique | top_counts |
|---------------|-----------|---------------|---------|----------|---------------------------------|
| gender | 0 | 1 | FALSE | 2 | 0: 1867, 1: 1736 |
| race | 0 | 1 | FALSE | 4 | 1: 2350, 3: 720, 4: 358, 2: 175 |
| smoking | 0 | 1 | FALSE | 3 | 0: 2176, 1: 1072, 2: 355 |
| hypertension | 0 | 1 | FALSE | 2 | 0: 1886, 1: 1717 |
| diabetes | 0 | 1 | FALSE | 2 | 0: 3065, 1: 538 |
| vaccine | 0 | 1 | FALSE | 2 | 1: 2119, 0: 1484 |
| severity | 0 | 1 | FALSE | 2 | 0: 3252, 1: 351 |
| study | 0 | 1 | FALSE | 3 | B: 2201, A: 718, C: 684 |
| length_ind | 0 | 1 | FALSE | 2 | yes: 2491, no: 1112 |

Variable type: numeric

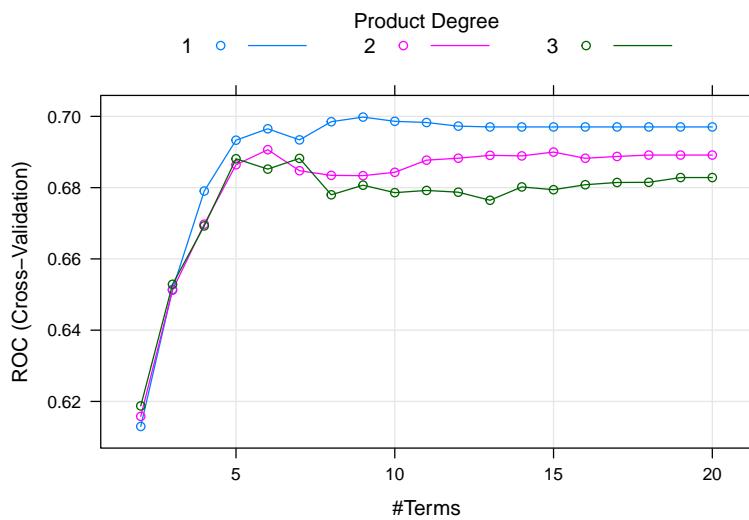
| skim_variable | n_missing | complete_rate | mean | sd | p0 | p25 | p50 | p75 | p100 | hist |
|---------------|-----------|---------------|---------|---------|-------|--------|--------|--------|--------|------|
| id | 0 | 1 | 4979.49 | 2845.35 | 2.0 | 2528.0 | 4987.0 | 7438.0 | 9999.0 | |
| age | 0 | 1 | 60.10 | 4.46 | 45.0 | 57.0 | 60.0 | 63.0 | 79.0 | |
| height | 0 | 1 | 170.02 | 5.93 | 148.1 | 166.1 | 170.0 | 173.9 | 189.1 | |
| weight | 0 | 1 | 79.91 | 7.08 | 57.1 | 75.0 | 79.9 | 84.7 | 104.2 | |
| bmi | 0 | 1 | 27.71 | 2.77 | 19.6 | 25.8 | 27.6 | 29.4 | 38.4 | |
| SBP | 0 | 1 | 130.01 | 7.88 | 103.0 | 125.0 | 130.0 | 135.0 | 156.0 | |
| LDL | 0 | 1 | 110.51 | 19.83 | 32.0 | 97.0 | 111.0 | 125.0 | 173.0 | |
| recovery_time | 0 | 1 | 42.98 | 29.46 | 2.0 | 28.0 | 39.0 | 50.0 | 365.0 | |



Penalized logistic regression

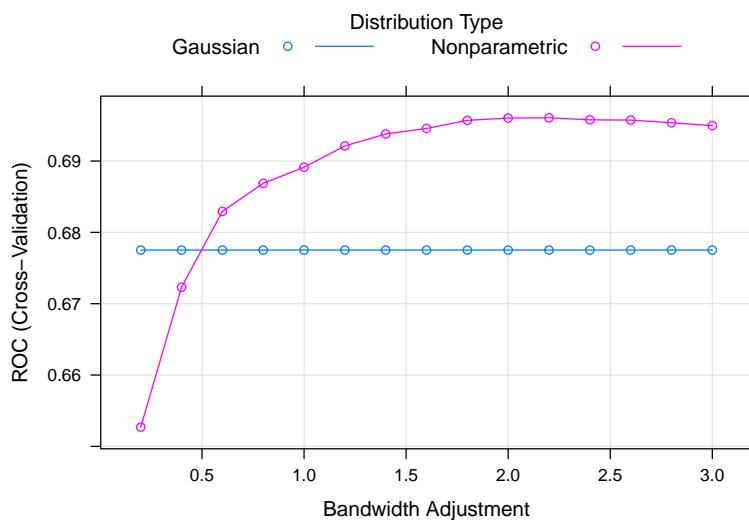


Multivariate adaptive regression splines classifier (MARS)

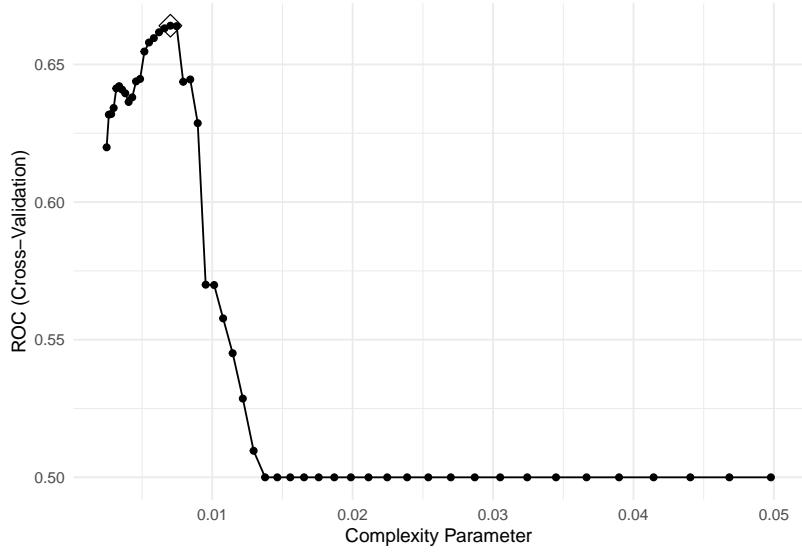


Quadratic discriminant analysis (QDA)

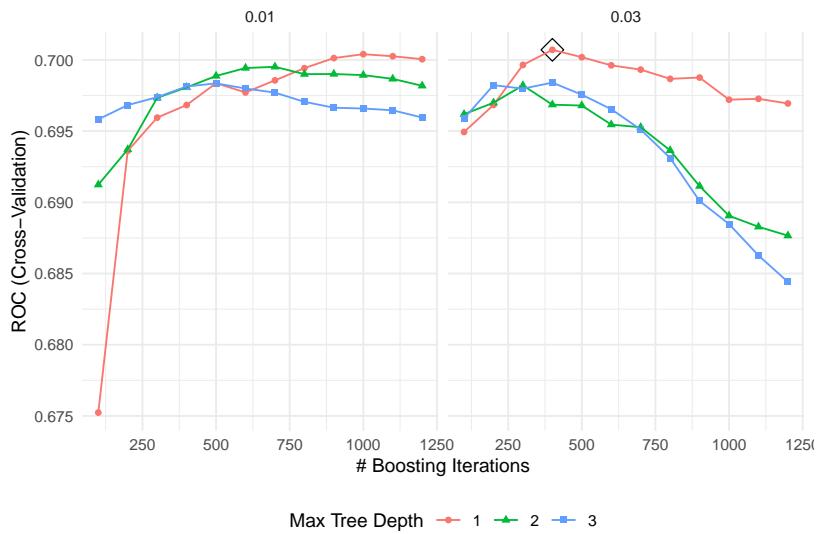
Naive Bayes Classifier



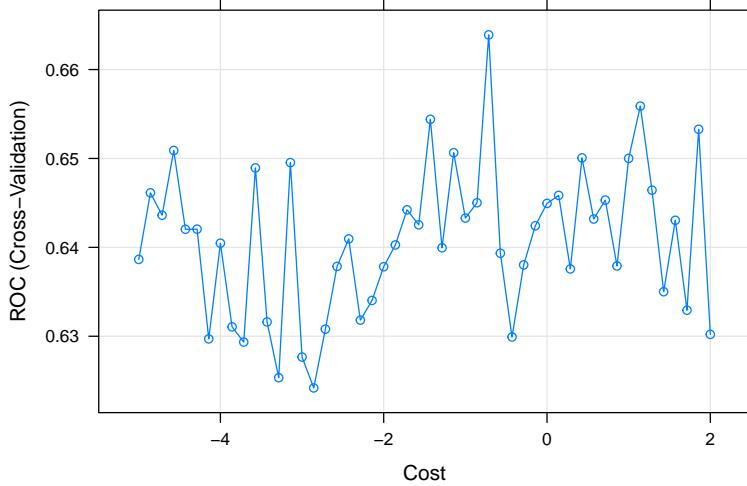
Classification Tree



Adaptive boosting classifier



Support vector classifier with linear kernel



Support vector classifier with radial kernel

