Covid Recovery Analysis - Secondary Analysis

Exploratory analysis for classification

Table 1: Data summary
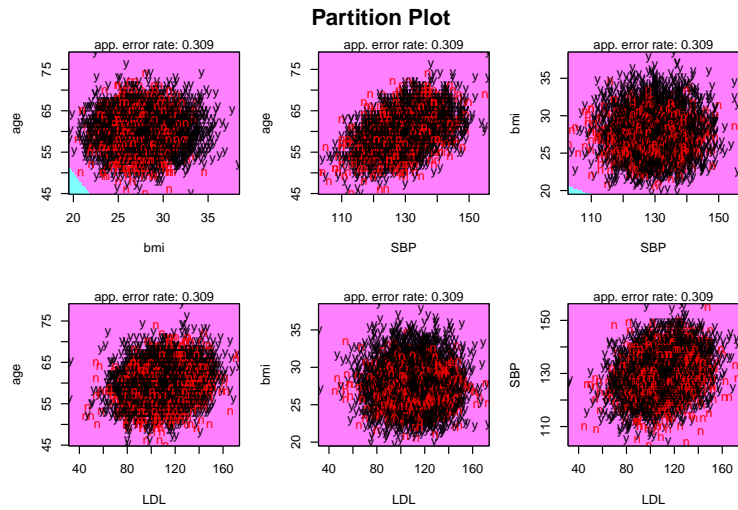
Name	Piped data
Number of rows	3603
Number of columns	17
Column type frequency:	
factor	9
numeric	8
Group variables	None

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	

Partition Plot

app. error rate: 0.309

Penalized logistic regression

Regularization Parameter

0.000372699966223616		0.00305959206434424	
0.00056783242423576		0.00466148574327131	
0.000865129303016903		0.00710207402743375	
0.00131808026275682		0.0108204676081991	
0.00200818024890684		0.0164856799306543	
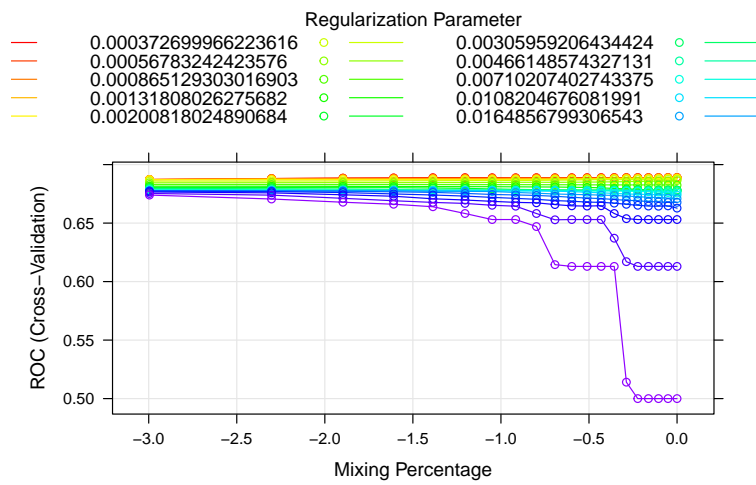
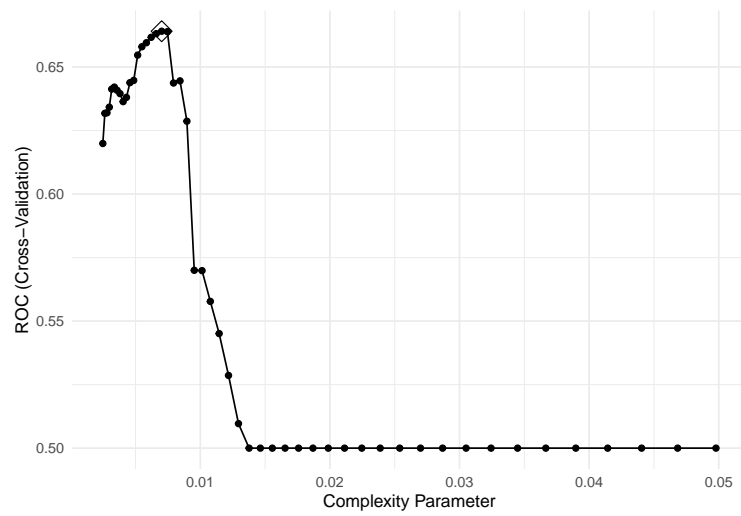# Multivariate adaptive regression splines classifier (MARS)
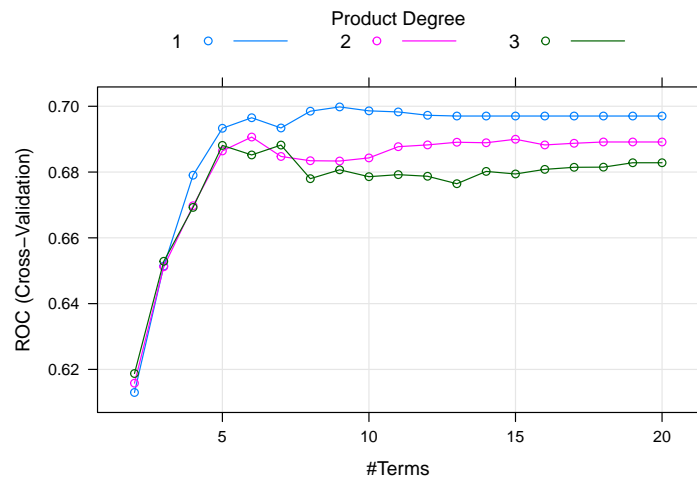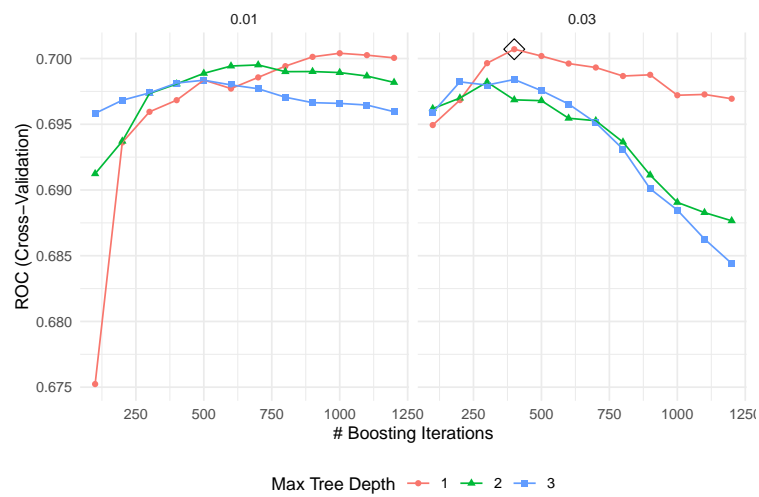

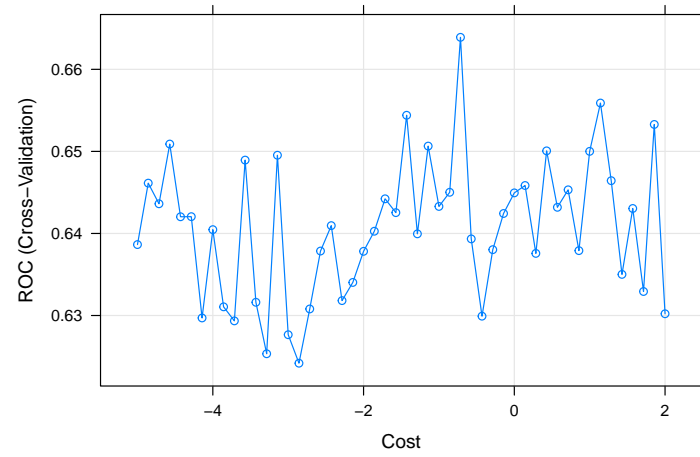# Quadratic discriminant analysis (QDA)

Naive Bayes Classifer
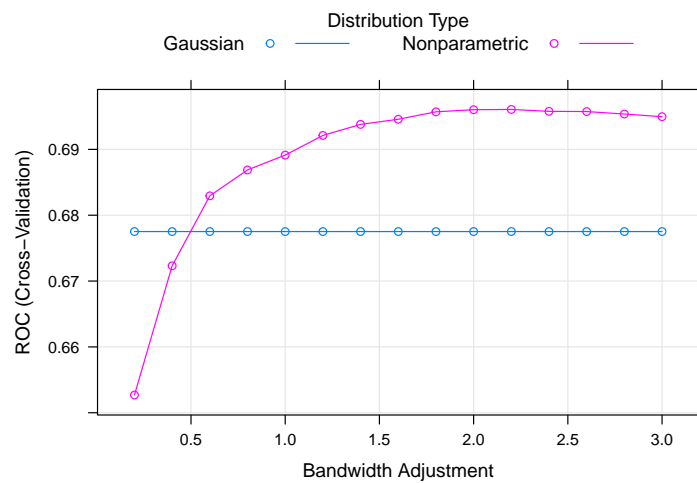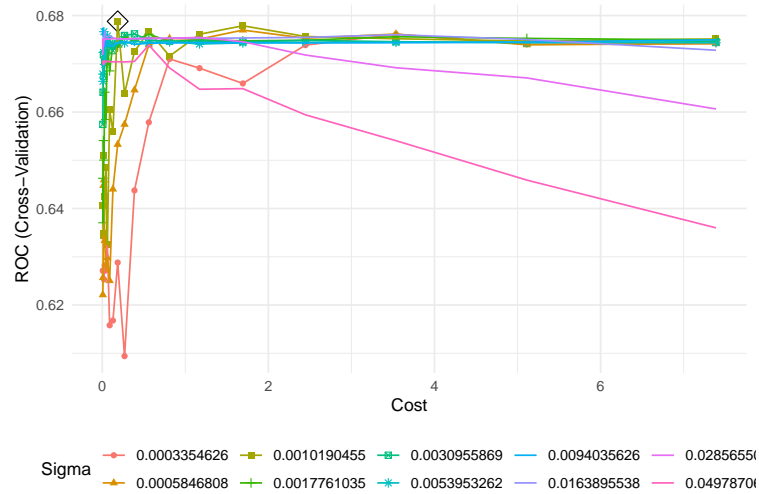
Classification Tree

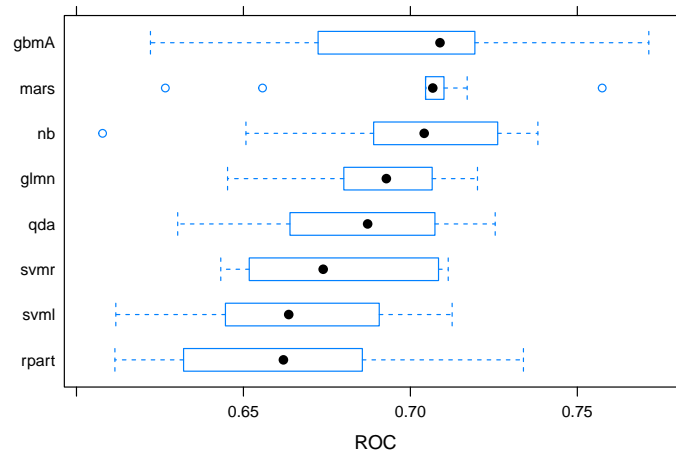## Adaptive boosting classfier

Support vector classifier with linear kernel

Support vector classifier with radial kernel
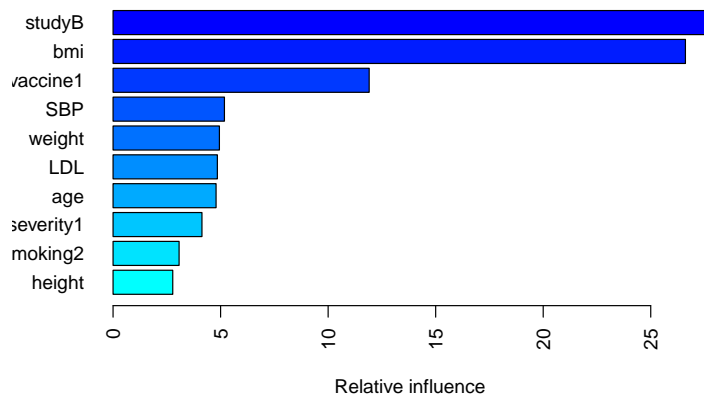
Models Comparsion based on Area Under Curves

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

	rel.inf
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 intepretation

Partial dependence plot by BMI

Partial dependence plot by SBP and LDL

ICE plot for bmi versus probability of long recovery (>30) in logit