

Classification of Acute Respiratory Distress Syndrome

Robert Edwards

Data Preparation

Scaling the continuous variables

Dropping NAs from dataset

Dropping variables with >40% NAs (PreECMO_Albumin)

Train / Validation / Test

Call:

```
glm(formula = ECMO_Survival ~ . - Pt_ID, family = binomial(link = "logit"),  
    data = train.ARDS)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.36813	0.01364	0.38692	0.69421	1.70684

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	2.06639	0.61058	3.384	0.000714	***
Genderw	-0.06459	0.56055	-0.115	0.908268	
Indication	-0.09829	0.15104	-0.651	0.515206	
Duration_ECMO	-0.17506	0.25882	-0.676	0.498793	
Age	-0.33419	0.29714	-1.125	0.260712	
PreECMO_RR	-0.49082	0.24983	-1.965	0.049461	*
PreECMO_Vt	0.01795	0.26295	0.068	0.945568	
PreECMO_FiO2	0.52923	0.52941	1.000	0.317471	
PreECMO_Ppeak	-0.26456	0.35289	-0.750	0.453442	
PreECMO_Pmean	0.43753	0.48817	0.896	0.370111	
PreECMO_PEEP	0.20183	0.42513	0.475	0.634966	
PreECMO_PF	1.49079	0.83594	1.783	0.074527	.
PreECMO_SpO2	-0.91677	0.47902	-1.914	0.055640	.
PreECMO_PaCO2	0.44623	0.63171	0.706	0.479953	
PreECMO_pH	0.89219	0.79236	1.126	0.260171	
PreECMO_BE	-0.68108	0.65022	-1.047	0.294890	
PreECMO_Lactate	-0.61473	0.39306	-1.564	0.117827	
PreECMO_NAdose	-0.11791	0.25316	-0.466	0.641392	
PreECMO_MAP	0.07270	0.29291	0.248	0.803977	
PreECMO_Creatinine	0.02676	0.28414	0.094	0.924962	
PreECMO_Urea	-0.22805	0.34557	-0.660	0.509295	

```

PreECMO_CK      0.08294    0.24158    0.343 0.731348
PreECMO_Bilirubin 0.82880    0.55468    1.494 0.135127
PreECMO_CRP     -0.28476    0.35144   -0.810 0.417778
PreECMO_Fibrinogen -0.20003    0.31744   -0.630 0.528603
PreECMO_Ddimer  -0.13960    0.29423   -0.474 0.635175
PreECMO_ATIII   -0.04830    0.17355   -0.278 0.780781
PreECMO_Leukocytes 0.04779    0.24911    0.192 0.847868
PreECMO_Platelets 0.45217    0.35559    1.272 0.203515
PreECMO_TNFa     1.69276    1.10178    1.536 0.124443
PreECMO_IL6     -1.39933    0.75665   -1.849 0.064405 .
PreECMO_IL8      0.84094    0.46438    1.811 0.070158 .
PreECMO_siIL2   -0.52497    0.38638   -1.359 0.174240
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

(Dispersion parameter for binomial family taken to be 1)

```

Null deviance: 170.92 on 155 degrees of freedom
Residual deviance: 130.32 on 123 degrees of freedom
AIC: 196.32

```

Number of Fisher Scoring iterations: 7

A quick note about the `plogis` function: The `glm()` procedure with `family="binomial"` will build the logistic regression model on the given formula. When we use the `predict` function on this model, it will predict the $\log(\text{odds})$ of the Y variable. This is not what we ultimately want because, the predicted values may not lie within the 0 and 1 range as expected. So, to convert it into prediction probability scores that is bound between 0 and 1, we use the `plogis()`. For more info see (blog on logistic regression)[<http://r-statistics.co/Logistic-Regression-With-R.html>].

Optimal Prediction Probability Cutoff for the Model

The default cutoff prediction probability score is 0.5 or the ratio of — in the training data. But sometimes, tuning the probability cutoff can improve the accuracy in both the development and validation samples. The `InformationValue::optimalCutoff` function provides ways to find the optimal cutoff to improve the prediction of —, —, both — and — and o reduce the misclassification error. Let's compute the optimal score that minimizes the misclassification error for the above model.

Model Diagnostics

Logistic Regression LASSO

```
[1] "Iteration 1 of 11"  
[1] "Iteration 2 of 11"  
[1] "Iteration 3 of 11"  
[1] "Iteration 4 of 11"  
[1] "Iteration 5 of 11"  
[1] "Iteration 6 of 11"  
[1] "Iteration 7 of 11"  
[1] "Iteration 8 of 11"  
[1] "Iteration 9 of 11"  
[1] "Iteration 10 of 11"  
[1] "Iteration 11 of 11"
```

	nets.a	nets.mse
[1,]	0.0	0
[2,]	0.1	0
[3,]	0.2	0
[4,]	0.3	0
[5,]	0.4	0
[6,]	0.5	0
[7,]	0.6	0
[8,]	0.7	0
[9,]	0.8	0
[10,]	0.9	0
[11,]	1.0	0

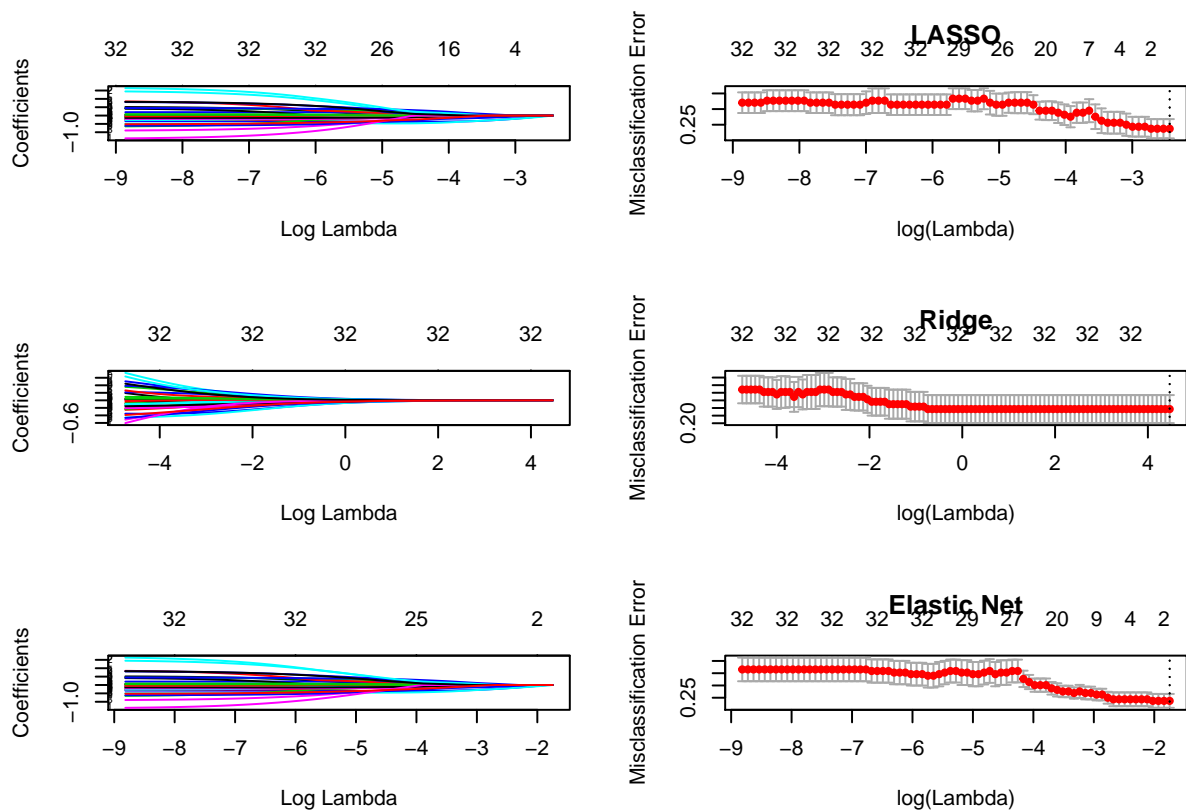


Figure 1: Lasso, Ridge, Elastic Net shrinkage and lambda cross-validation.

34 x 1 sparse Matrix of class "dgCMatrix"

```

      1
(Intercept) 1.168206
(Intercept) .
Genderw      .
Indication   .
Duration_ECMO .
Age          .
PreECMO_RR   .
PreECMO_Vt   .
PreECMO_FiO2 .
PreECMO_Ppeak .
PreECMO_Pmean .
PreECMO_PEEP .
PreECMO_PF   .
PreECMO_SpO2 .
PreECMO_PaCO2 .
PreECMO_pH   .
PreECMO_BE   .
PreECMO_Lactate .
PreECMO_NAdose .
PreECMO_MAP   .
PreECMO_Creatinine .

```

```

PreECMO_Urea      .
PreECMO_CK        .
PreECMO_Bilirubin .
PreECMO_CRP       .
PreECMO_Fibrinogen .
PreECMO_Ddimer    .
PreECMO_ATIII     .
PreECMO_Leukocytes .
PreECMO_Platelets .
PreECMO_TNFa      .
PreECMO_IL6       .
PreECMO_IL8       .
PreECMO_siIL2     .

```

Model Diagnostics

```

pred.valid.lasso
1
N 18
Y 60

```

K-Nearest Neighbors

CVA

LDA

Call:

```
lda(ECMO_Survival ~ ., data = train.ARDS)
```

Prior probabilities of groups:

	N	Y
	0.2371795	0.7628205

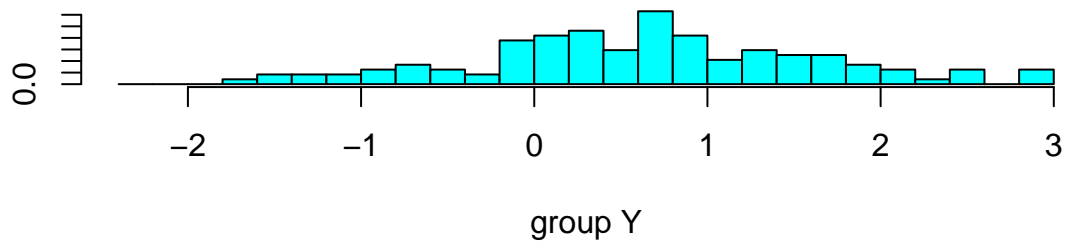
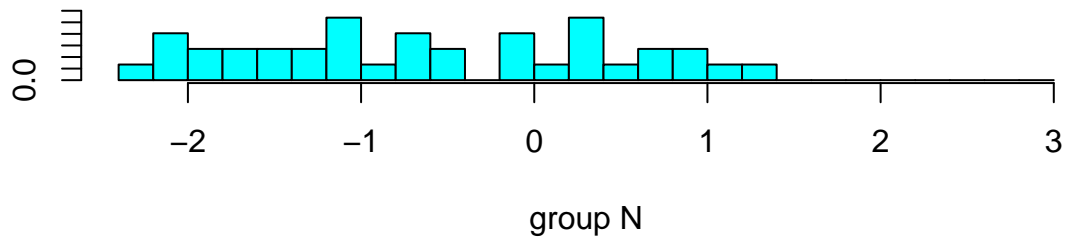
Group means:

	Pt_ID	Genderw	Indication	Duration_ECMO	Age	PreECMO_RR
N	1184.595	0.3513514	3.135135	0.17011326	0.49130348	0.3613807
Y	1306.504	0.3025210	3.025210	0.07056348	0.04243535	-0.1421564
	PreECMO_Vt	PreECMO_FiO2	PreECMO_Ppeak	PreECMO_Pmean	PreECMO_PEEP	
N	-0.14492842	0.1509813	0.04812984	-0.06664071	-0.07057527	
Y	0.03175897	0.0658993	-0.14141295	-0.02782580	-0.03489447	
	PreECMO_PF	PreECMO_SpO2	PreECMO_PaCO2	PreECMO_pH	PreECMO_BE	
N	-0.15494185	0.045583964	0.04865712	-0.22205686	-0.08825413	
Y	-0.01992909	0.008969737	-0.08852880	0.02194501	-0.05171488	
	PreECMO_Lactate	PreECMO_NAdose	PreECMO_MAP	PreECMO_Creatinine		
N	0.25283223	0.07537849	0.02183526	0.18529315		
Y	-0.06990024	-0.06305472	0.09965918	0.09892448		
	PreECMO_Urea	PreECMO_CK	PreECMO_Bilirubin	PreECMO_CRP	PreECMO_Fibrinogen	
N	0.157008949	0.02465349	-0.07508043	-0.04183654	0.0009114089	
Y	0.005798352	0.09209170	0.05693255	-0.05927163	-0.0642007345	
	PreECMO_Ddimer	PreECMO_ATIII	PreECMO_Leukocytes	PreECMO_Platelets		
N	0.06401838	-0.05630207	-0.07322120	-0.18851944		
Y	-0.10890629	0.02885875	-0.02137385	0.09582999		
	PreECMO_TNFa	PreECMO_IL6	PreECMO_IL8	PreECMO_siIL2		
N	-0.12734461	-0.01846578	-0.01379272	0.14233997		
Y	0.08328052	-0.06216531	0.09433882	-0.02785839		

Coefficients of linear discriminants:

	LD1
Pt_ID	0.000835365
Genderw	-0.074231374
Indication	-0.101733445
Duration_ECMO	-0.002499184
Age	-0.321287409
PreECMO_RR	-0.411075151
PreECMO_Vt	0.054459987
PreECMO_FiO2	0.172231162
PreECMO_Ppeak	-0.297988456
PreECMO_Pmean	0.343222196
PreECMO_PEEP	0.269872473
PreECMO_PF	0.715844394
PreECMO_SpO2	-0.587180764
PreECMO_PaCO2	0.086914800
PreECMO_pH	0.395770223
PreECMO_BE	-0.356129003
PreECMO_Lactate	-0.366484335

PreECMO_NAdose	-0.084296551
PreECMO_MAP	0.013741071
PreECMO_Creatinine	0.053123640
PreECMO_Urea	-0.226439830
PreECMO_CK	0.076304534
PreECMO_Bilirubin	0.409377382
PreECMO_CRP	-0.169903883
PreECMO_Fibrinogen	-0.074256637
PreECMO_Ddimer	-0.086370143
PreECMO_ATIII	-0.022278904
PreECMO_Leukocytes	0.115636210
PreECMO_Platelets	0.233476858
PreECMO_TNFa	0.277609647
PreECMO_IL6	-0.700329140
PreECMO_IL8	0.512021236
PreECMO_siIL2	-0.457549093



```
[1] "class"      "posterior" "x"
```

	N	Y
N	3	15
Y	8	52

	LD1
Pt_ID	0.000835365
Genderw	-0.074231374
Indication	-0.101733445
Duration_ECMO	-0.002499184
Age	-0.321287409
PreECMO_RR	-0.411075151
PreECMO_Vt	0.054459987

PreECMO_FiO2	0.172231162
PreECMO_Ppeak	-0.297988456
PreECMO_Pmean	0.343222196
PreECMO_PEEP	0.269872473
PreECMO_PF	0.715844394
PreECMO_SpO2	-0.587180764
PreECMO_PaCO2	0.086914800
PreECMO_pH	0.395770223
PreECMO_BE	-0.356129003
PreECMO_Lactate	-0.366484335
PreECMO_NAdose	-0.084296551
PreECMO_MAP	0.013741071
PreECMO_Creatinine	0.053123640
PreECMO_Urea	-0.226439830
PreECMO_CK	0.076304534
PreECMO_Bilirubin	0.409377382
PreECMO_CRP	-0.169903883
PreECMO_Fibrinogen	-0.074256637
PreECMO_Ddimer	-0.086370143
PreECMO_ATIII	-0.022278904
PreECMO_Leukocytes	0.115636210
PreECMO_Platelets	0.233476858
PreECMO_TNFa	0.277609647
PreECMO_IL6	-0.700329140
PreECMO_IL8	0.512021236
PreECMO_siIL2	-0.457549093

QDA

Call:

```
qda(ECMO_Survival ~ ., data = train.ARDS)
```

Prior probabilities of groups:

	N	Y
	0.2371795	0.7628205

Group means:

	Pt_ID	Genderw	Indication	Duration_ECMO	Age	PreECMO_RR
N	1184.595	0.3513514	3.135135	0.17011326	0.49130348	0.3613807
Y	1306.504	0.3025210	3.025210	0.07056348	0.04243535	-0.1421564
	PreECMO_Vt	PreECMO_FiO2	PreECMO_Ppeak	PreECMO_Pmean	PreECMO_PEEP	
N	-0.14492842	0.1509813	0.04812984	-0.06664071	-0.07057527	
Y	0.03175897	0.0658993	-0.14141295	-0.02782580	-0.03489447	
	PreECMO_PF	PreECMO_SpO2	PreECMO_PaCO2	PreECMO_pH	PreECMO_BE	
N	-0.15494185	0.045583964	0.04865712	-0.22205686	-0.08825413	
Y	-0.01992909	0.008969737	-0.08852880	0.02194501	-0.05171488	
	PreECMO_Lactate	PreECMO_NAdose	PreECMO_MAP	PreECMO_Creatinine		
N	0.25283223	0.07537849	0.02183526	0.18529315		
Y	-0.06990024	-0.06305472	0.09965918	0.09892448		
	PreECMO_Urea	PreECMO_CK	PreECMO_Bilirubin	PreECMO_CRP	PreECMO_Fibrinogen	
N	0.157008949	0.02465349	-0.07508043	-0.04183654	0.0009114089	
Y	0.005798352	0.09209170	0.05693255	-0.05927163	-0.0642007345	
	PreECMO_Ddimer	PreECMO_ATIII	PreECMO_Leukocytes	PreECMO_Platelets		
N	0.06401838	-0.05630207	-0.07322120	-0.18851944		
Y	-0.10890629	0.02885875	-0.02137385	0.09582999		
	PreECMO_TNFa	PreECMO_IL6	PreECMO_IL8	PreECMO_siIL2		
N	-0.12734461	-0.01846578	-0.01379272	0.14233997		
Y	0.08328052	-0.06216531	0.09433882	-0.02785839		

```
[1] "class"      "posterior"
```

	N	Y
N	0	18
Y	0	60

	N	Y
81	1.579597e-172	1
101	8.463081e-137	1
247	1.355908e-78	1
112	7.117467e-242	1
379	1.879201e-69	1
87	9.721819e-88	1
438	3.783116e-278	1
349	0.000000e+00	1
108	0.000000e+00	1
341	9.703442e-121	1
353	4.290646e-90	1
344	1.855858e-218	1
442	4.619805e-314	1

355	4.639690e-86	1
315	2.905056e-38	1
240	0.000000e+00	1
260	2.788294e-92	1
357	0.000000e+00	1

NULL

Trees

Random Forest

Call:

```
randomForest(formula = ECMO_Survival ~ ., data = train.ARDS[,      -1], proximity = TRUE, importance = FALSE,
              Type of random forest: classification
              Number of trees: 1000
```

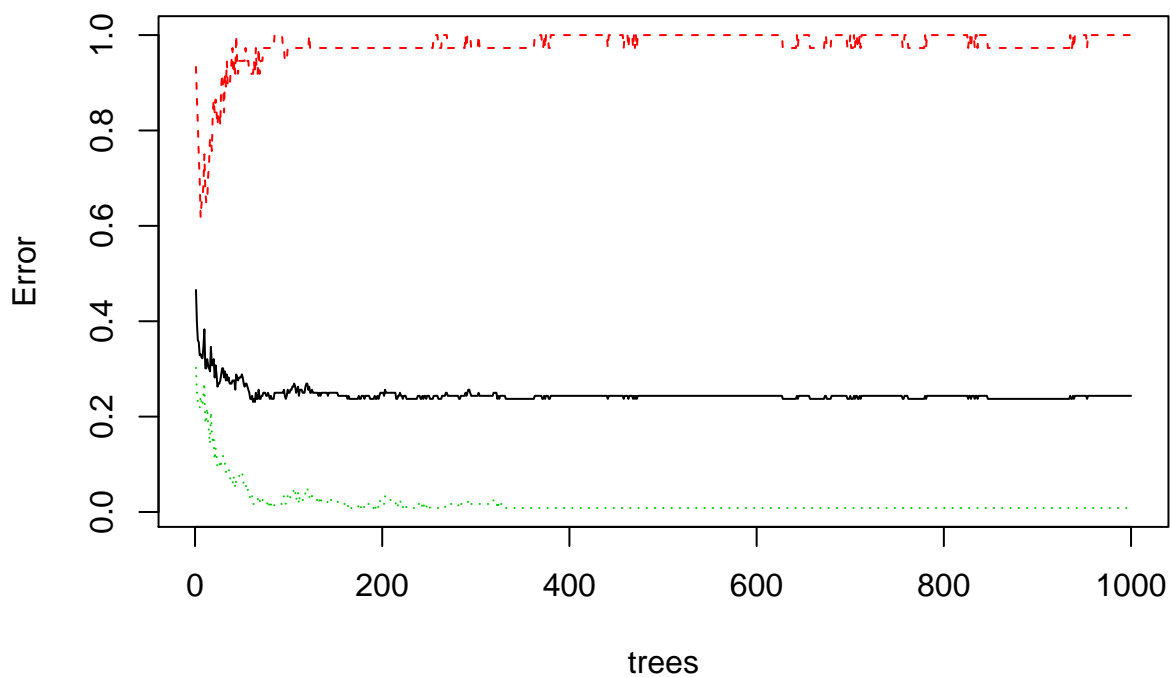
No. of variables tried at each split: 5

OOB estimate of error rate: 24.36%

Confusion matrix:

```
      N   Y class.error
N 0   37 1.0000000000
Y 1  118 0.008403361
```

model.rf



```
pred.valid.rf
  N  Y
1  1 17
2  0 60
```

Support Vector Machines

Call:

```
svm(formula = ECMO_Survival ~ ., data = train.ARDS[, -1], kernel = "polynomial",  
     cost = 10, decision.values = TRUE)
```

Parameters:

```
SVM-Type: C-classification  
SVM-Kernel: polynomial  
cost: 10  
degree: 3  
gamma: 0.03030303  
coef.0: 0
```

Number of Support Vectors: 135

```
pred.valid.svm  
  N  Y  
N  1 17  
Y  2 58
```

Model Selection

Table 1: Accuracy on validation set

	Accuracy	True Positive Rate	False Positive Rate
Logistic	0.744	0.400	0.100
KNN	0.731	0.056	0.233
LDA	0.705	0.273	0.133
QDA	0.769	NaN	0.000
RF	0.782	1.000	0.000
SVM	0.756	0.333	0.033

ROC

