

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
-2.36703	0.08456	0.32105	0.59363	1.58562

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	2.1574702	0.6753678	3.195	0.001401	**
Genderw	-0.2041289	0.6587942	-0.310	0.756673	
Indication	-0.0267857	0.1478585	-0.181	0.856244	
Duration_ECMO	-0.4838630	0.2432947	-1.989	0.046724	*
Age	-1.3539978	0.3823702	-3.541	0.000399	***
PreECMO_RR	-0.4362082	0.3118592	-1.399	0.161893	
PreECMO_Vt	-0.0005162	0.2829431	-0.002	0.998544	
PreECMO_FiO2	0.1508517	0.5521989	0.273	0.784712	
PreECMO_Ppeak	0.8563842	0.5205988	1.645	0.099970	.
PreECMO_Pmean	-1.6602896	0.7104315	-2.337	0.019438	*
PreECMO_PEEP	1.4221454	0.5359030	2.654	0.007961	**
PreECMO_PF	-0.0069376	0.5675699	-0.012	0.990247	
PreECMO_SpO2	0.2939924	0.4369905	0.673	0.501096	
PreECMO_PaCO2	-2.3747024	0.9494229	-2.501	0.012377	*
PreECMO_pH	-2.4280431	1.0424525	-2.329	0.019850	*
PreECMO_BE	1.2967913	0.7571707	1.713	0.086771	.
PreECMO_Lactate	-1.0312345	0.4435300	-2.325	0.020069	*
PreECMO_NAdose	0.2122004	0.2533062	0.838	0.402186	
PreECMO_MAP	0.1265281	0.2843933	0.445	0.656388	
PreECMO_Creatinine	-0.0349847	0.3744639	-0.093	0.925565	
PreECMO_Urea	-0.0783679	0.4637719	-0.169	0.865813	

```

PreECMO_CK          -0.7599354  0.3274305  -2.321  0.020292 *
PreECMO_Bilirubin   -0.6097946  0.3023562  -2.017  0.043715 *
PreECMO_CRP          -0.3027948  0.3653298  -0.829  0.407203
PreECMO_Fibrinogen  -0.4354804  0.3777785  -1.153  0.249017
PreECMO_Ddimer       0.0429057  0.3155602   0.136  0.891847
PreECMO_ATIII       -0.3198426  0.3393163  -0.943  0.345881
PreECMO_Leukocytes  0.3570072  0.4007024   0.891  0.372954
PreECMO_Platelets    0.0426731  0.3074918   0.139  0.889626
PreECMO_TNFa        -0.0896648  0.2578221  -0.348  0.728007
PreECMO_IL6         -0.3885338  0.5177852  -0.750  0.453028
PreECMO_IL8         -0.5832561  1.0844609  -0.538  0.590694
PreECMO_siIL2        0.0791202  0.3923069   0.202  0.840167
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

(Dispersion parameter for binomial family taken to be 1)

```

Null deviance: 155.12  on 154  degrees of freedom
Residual deviance: 112.86  on 122  degrees of freedom
(70 observations deleted due to missingness)
AIC: 178.86

```

Number of Fisher Scoring iterations: 6

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

```

[1] NA
      N      Y
0.2272727 0.8431373

```

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

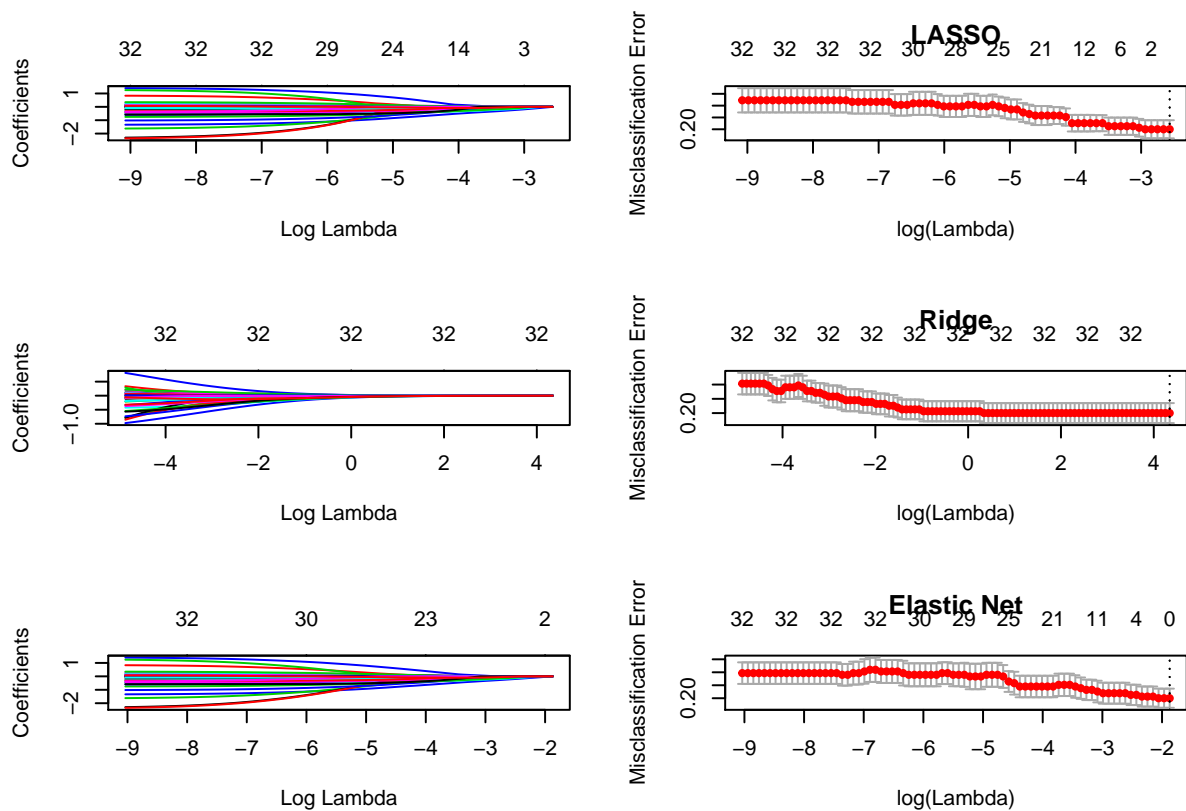


Figure 1: Lasso, Ridge, Elastic Net shrinkage and lambda cross-validation.

34 x 1 sparse Matrix of class "dgCMatrix"

```

      1
(Intercept) 1.386294
(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

```
[1] 0
```

	N	Y
	1.00	0.25

K-Nearest Neighbors

LDA/QDA

Support Vector Machines