IDS 702

Logistic Regression Diagnostics for Predictions

Assessing predicted outcomes

Confusion matrix:

		Observed	
		Y=1	Y=0
Predicted	Y=1	TP (True Positives)	FP (False Positives)
	Y=0	FN (False Negatives)	TN (True Negatives)

Confusion matrix

		Observed	
		Y=1	Y=0
Predicted	Y=1	TP (True Positives)	FP (False Positives)
	Y=0	FN (False Negatives)	TN (True Negatives)

- True positive rate (sensitivity) = TP/(TP+FN)
- False positive rate (1-specificity) = FP/(FP+TN)
- True negative rate (specificity) = TN/(FP+TN)
- False negative rate = FN/(TP+FN)

Confusion matrix in R

```
> confusionMatrix(factor(ifelse(fitted(pumpmod)>0.5,"Ürgüp Sivrisi","Çerçevelik")),factor(pumpkin$Class),
positive="Ürgüp Sivrisi",mode="everything")
Confusion Matrix and Statistics

               Reference
Prediction      Çerçevelik Ürgüp Sivrisi
  Çerçevelik          1163           196
  Ürgüp Sivrisi        137          1004

               Accuracy : 0.8668
                 95% CI : (0.8529, 0.8799)
    No Information Rate : 0.52
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.7327

 Mcnemar's Test P-Value : 0.001481

            Sensitivity : 0.8367
            Specificity : 0.8946
         Pos Pred Value : 0.8799
         Neg Pred Value : 0.8558
              Precision : 0.8799
                 Recall : 0.8367
                     F1 : 0.8578
             Prevalence : 0.4800
         Detection Rate : 0.4016
   Detection Prevalence : 0.4564
      Balanced Accuracy : 0.8656

       'Positive' Class : Ürgüp Sivrisi
```

ROC Curves

- Ideally, we'd like high values of both sensitivity and specificity (low values of 1-specificity)
- The receiver operating characteristic (ROC) curve plots:
    - Sensitivity on Y-axis
    - 1-specificity on X-axis
- Evaluated at lots of different probability cutoff values
- Area Under the Curve (AUC) calculated as area under the ROC curve (what value do we want?)

ROC Curve in R

```
> library(pROC)
> roc(pumpkin$class.fac,fitted(pumpmod),print.thres=0.5,print.auc=T,plot=T,legacy.axes=T)
```