Desbalance de Clases

Machine Learning

Ignacio Sarmiento-Barbieri

Universidad de La Plata

Agenda

Classification

- We observe (y_i, X_i) $i = 1, \ldots, n$
- Probabilities
  - Logit
  - Random Forests
  - ...
  - Networks
- Classification

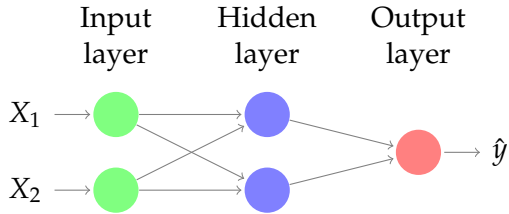
$$\hat{Y}_i = 1[\hat{p}_i > c] \tag{1}$$

Deep Learning: Recap

- Neural networks are linear combinations of inputs that are passed through nonlinear activation functions called nodes (or, in reference to the human brain, neurons).

Deep Learning: Recap

- Neural networks are linear combinations of inputs that are passed through nonlinear activation functions called nodes (or, in reference to the human brain, neurons).



Activation Functions

- Activation functions
  - $\text{ReLU}(x) = \max\{x, 0\}$.
  - $\text{sigmoid}(x) = \frac{1}{1+\exp(-x)}$.
  - $\tanh(x) = \frac{1-\exp(-2x)}{1+\exp(-2x)}$.
  - Others like: $cos(x)$, Radial basis function (RBF), Softplus, Hard tanh, etc ...
  - Radial basis function (RBF): $exp\left(\frac{1}{\sigma^2}||W-x||^2\right)$
  - Softplus: $log(1+e^x)$
  - Hard tanh: $max(-1, min(1, x))$
- Hidden unit design remains an active area of research, and many useful hidden unit types remain to be discovered

Output Functions

- The choice of cost function is tightly coupled with the choice of output unit.
- Most of the time, we simply use the distance between the data distribution and the model distribution.
  - Linear $y = \beta_0 + \sum_{k=1}^{K} \beta_k h_k \rightarrow \mathbb{R}$
  - Sigmoid (Logistic) $\rightarrow$ classification $\{0, 1\}$
  - Softmax $\rightarrow$ classification multiple categories

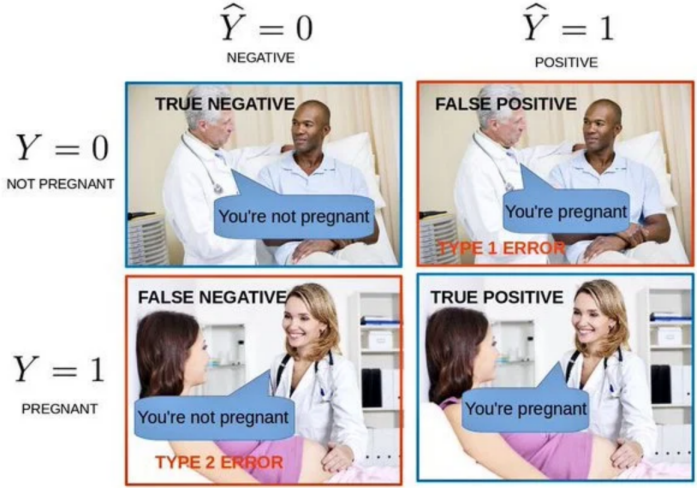
Example: Default

photo from https://www.dailydot.com/parsec/batman-1966-labels-tumblr-twitter-vine/

Agenda

1. Classification
2. Misclassification Rates
   - ROC curve
3. Desbalance de Clases

Misclassification Rates



Misclassification Rates

		$\hat{y}_i$	
		0	1
y_i	0	TN	FP
	1	FN	TP

- We have two types of error associated with this that we can use as a measure of performance

$$\begin{aligned} \textit{False Positive Rate} &= \frac{\textit{False Positives}}{\textit{Negatives}} \\ \textit{True Positive Rate} &= \frac{\textit{True Positives}}{\textit{Positives}} \end{aligned} \tag{2}$$

- Another names they receive:
  - False positive rate: Type I error, 1-Specificity
  - True positive rate: 1- Type II error, power, sensitivity.

Misclassification Rates

		$\hat{y}_i$	
		0	1
y_i	0	TN	FP
	1	FN	TP

- Another measures of performance using the predicted classes
  - Positive predicted values, also called: Precision, 1- false discovery proportion

$$Positive\ Predicted\ Values = \frac{True\ Positives}{PredictedPositives} \tag{3}$$

  - Negative predicted values

$$Negative\ Predicted\ Values = \frac{True\ Negatives}{PredictedNegatives} \tag{4}$$

Misclassification Rates

Accuracy

		$\hat{y}_i$	
		0	1
y_i	0	TN	FP
	1	FN	TP

- Accuracy: the fraction of predictions our model got right.

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \tag{5}$$

Misclassification Rates

F1 Score

		$\hat{y}_i$	
		0	1
y_i	0	TN	FP
	1	FN	TP

- The F1 Score is the harmonic mean of precision and recall. It is a way to combine both metrics into a single, useful metric.

$$F1 = 2\frac{Precision \times Recall}{Precision + Recall} \tag{6}$$

- The F1 score is particularly useful when you need to balance precision and recall, and there is an uneven class distribution (large number of actual negatives).

ROC

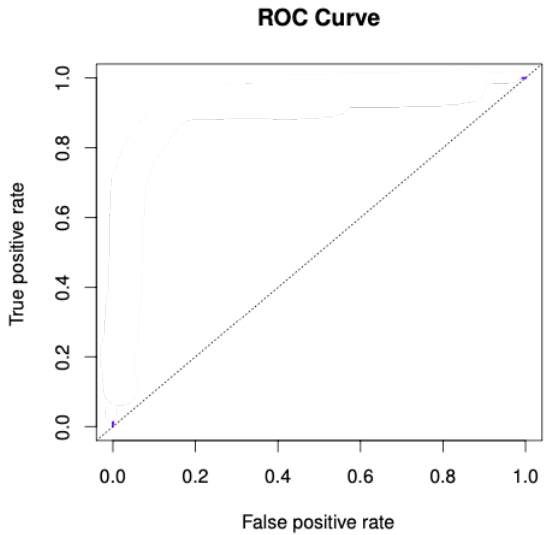
		$\hat{y}_i$ = 0	$\hat{y}_i$ = 1
y_i	0	TN	FP
	1	FN	TP

- A classification rule, or cutoff, is the probability p at which you predict
  - $\hat{y}_i = 0$ if $p_i < c$
  - $\hat{y}_i = 1$ if $p_i > c$
- Bayes classifier $c = 0.5$
- Changing c changes predictions, changes FP and FN
- There is a trade-off: reducing one error increases the other

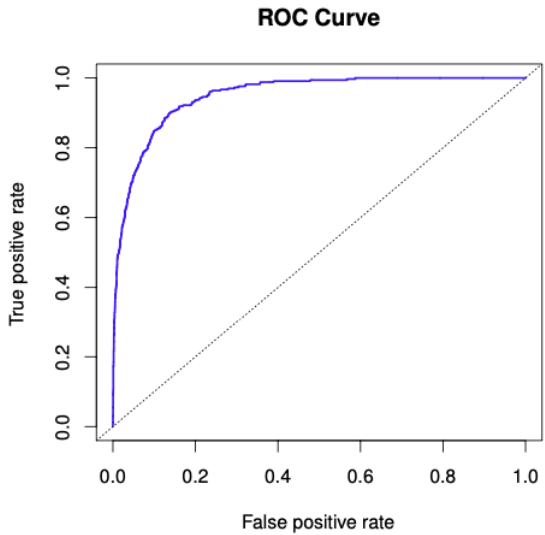
ROC

- ROC curve: Receiver operating characteristic curve
- ROC curve illustrates the trade-off of the classification rule
- Gives us the ability
  - Measure the predictive capacity of our model
  - Compare between models

ROC


ROC Curve
True positive rate
False positive rate
0.0
0.2
0.4
0.6
0.8
1.0

ROC


ROC Curve
True positive rate
False positive rate
0.0
0.2
0.4
0.6
0.8
1.0

Example: Default

photo from https://www.dailydot.com/parsec/batman-1966-labels-tumblr-twitter-vine/

Agenda

Desbalance de Clases: Motivation

- Al modelar clases discretas, las frecuencias relativas de las clases pueden tener un impacto significativo en la efectividad del modelo.
- Se produce un desequilibrio cuando una o más clases tienen proporciones muy bajas en los datos de entrenamiento en comparación con las otras clases.
- Ejemplos: Default, Pobreza, Tasa de Positividad de Covid, etc.
- Existen varias estrategias a utilizar y vamos a explorarla en el siguiente ejemplo

Example: Fraud

photo from https://www.dailydot.com/parsec/batman-1966-labels-tumblr-twitter-vine/