

Contents

- 1 Classification
 - Framework and notations

Contents

1 Classification

- Framework and notations
- Confusion matrix

Contents

1 Classification

- Framework and notations
- Confusion matrix
- ROC Curve

Contents

1 Classification

- Framework and notations
- Confusion matrix
- ROC Curve
- Logistic Regression motivation

Contents

- 1 Classification
 - Framework and notations
 - Confusion matrix
 - ROC Curve
 - Logistic Regression motivation
 - Logistic Regression algorithm

Contents

1 Classification

- Framework and notations
- Confusion matrix
- ROC Curve
- Logistic Regression motivation
- Logistic Regression algorithm
- Other approach K-Nearest Neighbours (KNN)

Contents

1 Classification

- Framework and notations
- Confusion matrix
- ROC Curve
- Logistic Regression motivation
- Logistic Regression algorithm
- Other approach K-Nearest Neighbours (KNN)
- Summary

Outline

1 Classification

- Framework and notations
 - Confusion matrix
 - ROC Curve
 - Logistic Regression motivation
 - Logistic Regression algorithm
 - Other approach K-Nearest Neighbours (KNN)
 - Summary

In fact, this is the same notation as in the linear models chapter, except that now Y_i is discrete and represent generally a class number, a qualitative variable $Y_i \in \{0, 1, \dots, K\}$, $Y_i \in \{\text{spam}, \text{ham}\}$

$$Y_i = h(X^i) = \beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \dots + \beta_p X_{i,p}$$

Moreover the prediction is a probability to be in the class k . In a two-class problem, you can decide Y_i is true if probability is > 0.5

Outline

1 Classification

- Framework and notations
- **Confusion matrix**
- ROC Curve
- Logistic Regression motivation
- Logistic Regression algorithm
- Other approach K-Nearest Neighbours (KNN)
- Summary

Vocabulary :

		Predicted condition		
Total population		Predicted Condition positive	Predicted Condition negative	Prevalence = $\frac{\sum \text{Condition positive}}{\sum \text{Total population}}$
True condition	condition positive	True positive	False Negative (Type II error)	True positive rate (TPR), Sensitivity, Recall $= \frac{\sum \text{True positive}}{\sum \text{Condition positive}}$
	condition negative	False Positive (Type I error)	True negative	False positive rate (FPR), Fall-out $= \frac{\sum \text{False positive}}{\sum \text{Condition negative}}$
Accuracy (ACC) = $\frac{\sum \text{True positive} + \sum \text{True negative}}{\sum \text{Total population}}$		Positive predictive value (PPV), Precision $= \frac{\sum \text{True positive}}{\sum \text{Test outcome positive}}$	False omission rate (FOR) $= \frac{\sum \text{False negative}}{\sum \text{Test outcome negative}}$	Positive likelihood ratio (LR+) = $\frac{\text{TPR}}{\text{FPR}}$
		False discovery rate (FDR) $= \frac{\sum \text{False positive}}{\sum \text{Test outcome positive}}$	Negative predictive value (NPV) $= \frac{\sum \text{True negative}}{\sum \text{Test outcome negative}}$	Negative likelihood ratio (LR-) = $\frac{\text{FNR}}{\text{TNR}}$

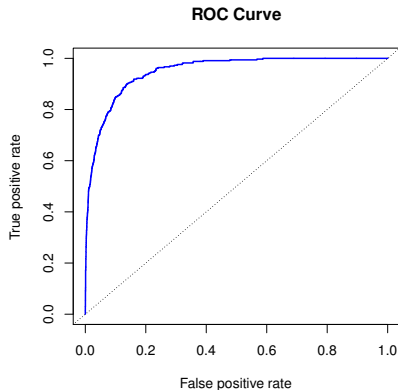
Beware : In the book Predicted is given horizontally
From wikipedia-en on ROC

Outline

1 Classification

- Framework and notations
- Confusion matrix
- ROC Curve
- Logistic Regression motivation
- Logistic Regression algorithm
- Other approach K-Nearest Neighbours (KNN)
- Summary

From communication theory Receiver Operator Curve :

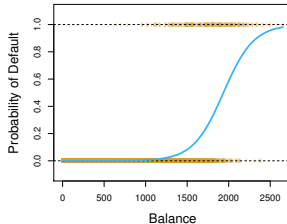
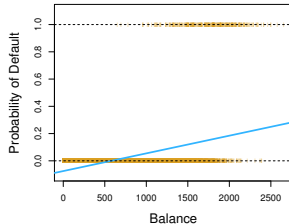


The associated measure is the Area Under the Curve (**AUC**), need to be near 1.

Outline

1 Classification

- Framework and notations
- Confusion matrix
- ROC Curve
- **Logistic Regression motivation**
- Logistic Regression algorithm
- Other approach K-Nearest Neighbours (KNN)
- Summary



- To avoid values outside $[0, 1]$, we use a specific function called **sigmoid** or **logistic function**
- $\text{sigmoid}(x) = \frac{1}{1+e^{-x}} = \frac{e^x}{e^x+1} \in [0, 1]$
- The Linear Regression is "filtered" by this function.

Outline

1 Classification

- Framework and notations
- Confusion matrix
- ROC Curve
- Logistic Regression motivation
- **Logistic Regression algorithm**
- Other approach K-Nearest Neighbours (KNN)
- Summary

- After some manipulation the model is :
 $\ln\left(\frac{p(X)}{1-p(X)}\right) = \beta_0 + \beta_1 X_1 + \epsilon$ where $p(X)$ is the probability that the response is 1
- Instead of using Least Square algorithm to adjust the β_i parameters, here we use a more statistical way, the Maximum Likelihood algorithm
- Obviously we can have multiple predictors :
 $\ln\left(\frac{p(X)}{1-p(X)}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_i X_p + \epsilon$

Outline

- 1 Classification
 - Framework and notations
 - Confusion matrix
 - ROC Curve
 - Logistic Regression motivation
 - Logistic Regression algorithm
 - **Other approach K-Nearest Neighbours (KNN)**
 - Summary

- In order to make a prediction for an observation $X = x$, the K training observations that are closest to x are identified.
- Then X is assigned to the class to which the plurality of these observations belong. Hence KNN is a completely **non-parametric** approach : no assumptions are made about the shape of the decision boundary.
- There is also a KNN Regression algorithm (book)

Outline

1 Classification

- Framework and notations
- Confusion matrix
- ROC Curve
- Logistic Regression motivation
- Logistic Regression algorithm
- Other approach K-Nearest Neighbours (KNN)
- Summary

- Prediction of probabilities ;
- Same minimization frameworks ;
- In general, all the algorithms have a Regression and a Classification behaviour.