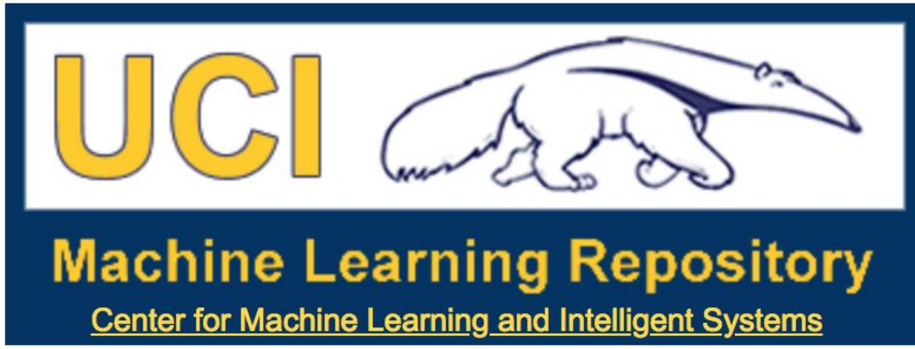

Tera

Aula #19

Algoritmos de Classificação: Regressão Logística

Gabriel Cypriano
28/ago/2018

Como será?



Bank Marketing Data Set

Download: Data Folder, Data Set Description

Abstract: The data is related with direct marketing campaigns (phone calls) of a Portuguese banking institution. The classification goal is to predict if the client will subscribe a term dep

Bank Marketing Data Set

Download: Data Folder, Data Set Description

Abstract: The data is related with direct marketing campaigns (phone calls) of a Portuguese banking institution. The classification goal is to predict if the client will subscribe a term dep

- Ligações de um banco português ofertando investimento financeiro
- Target: o cliente investiu ou não

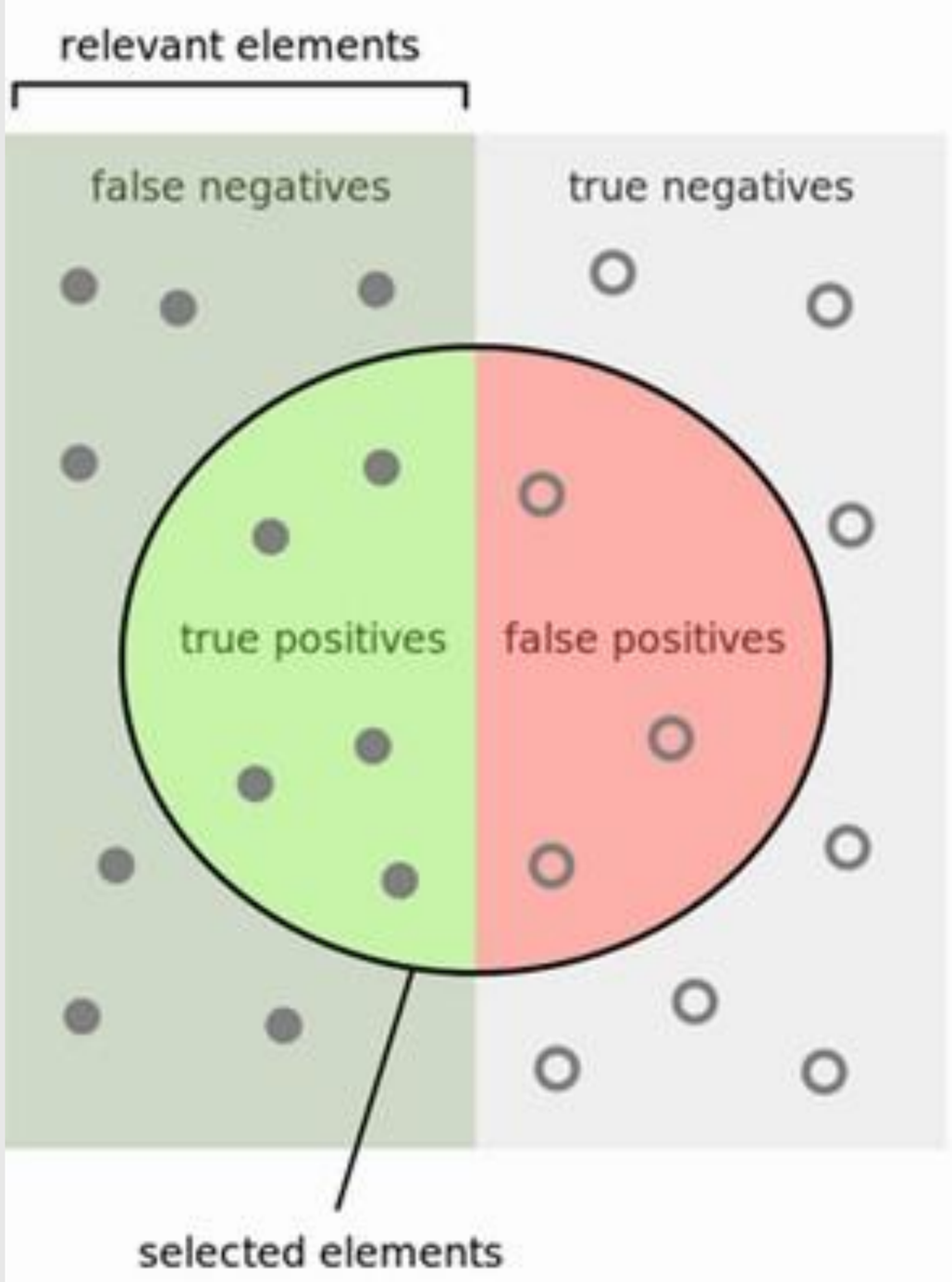
Análise Exploratória

Discussão

Qual métrica utilizar?

Accuracy Paradox

Predictive models with a given level of accuracy may have greater predictive power than models with higher accuracy.

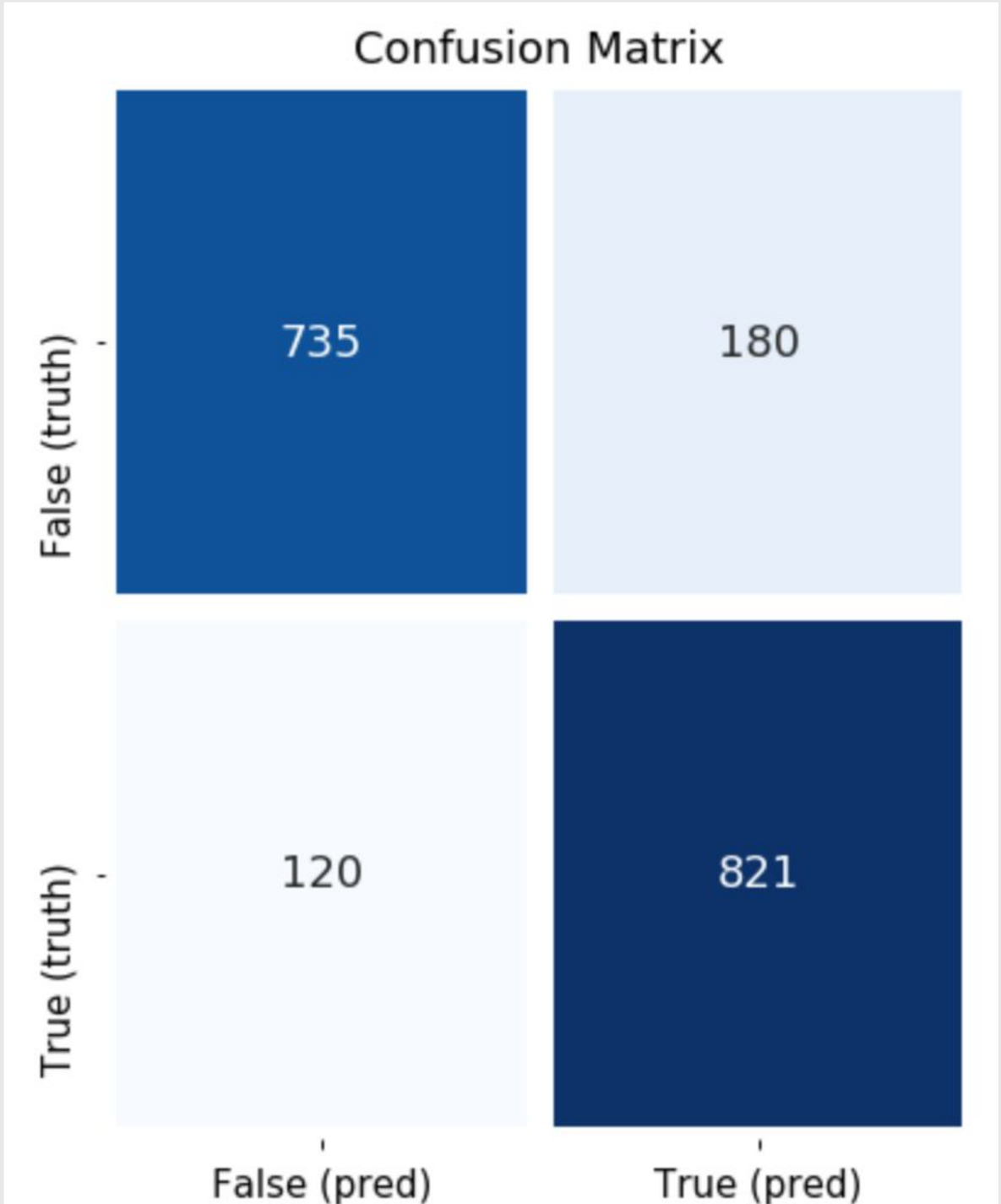


How many selected items are relevant?

$$\text{Precision} = \frac{\text{true positives}}{\text{true positives} + \text{false positives}}$$

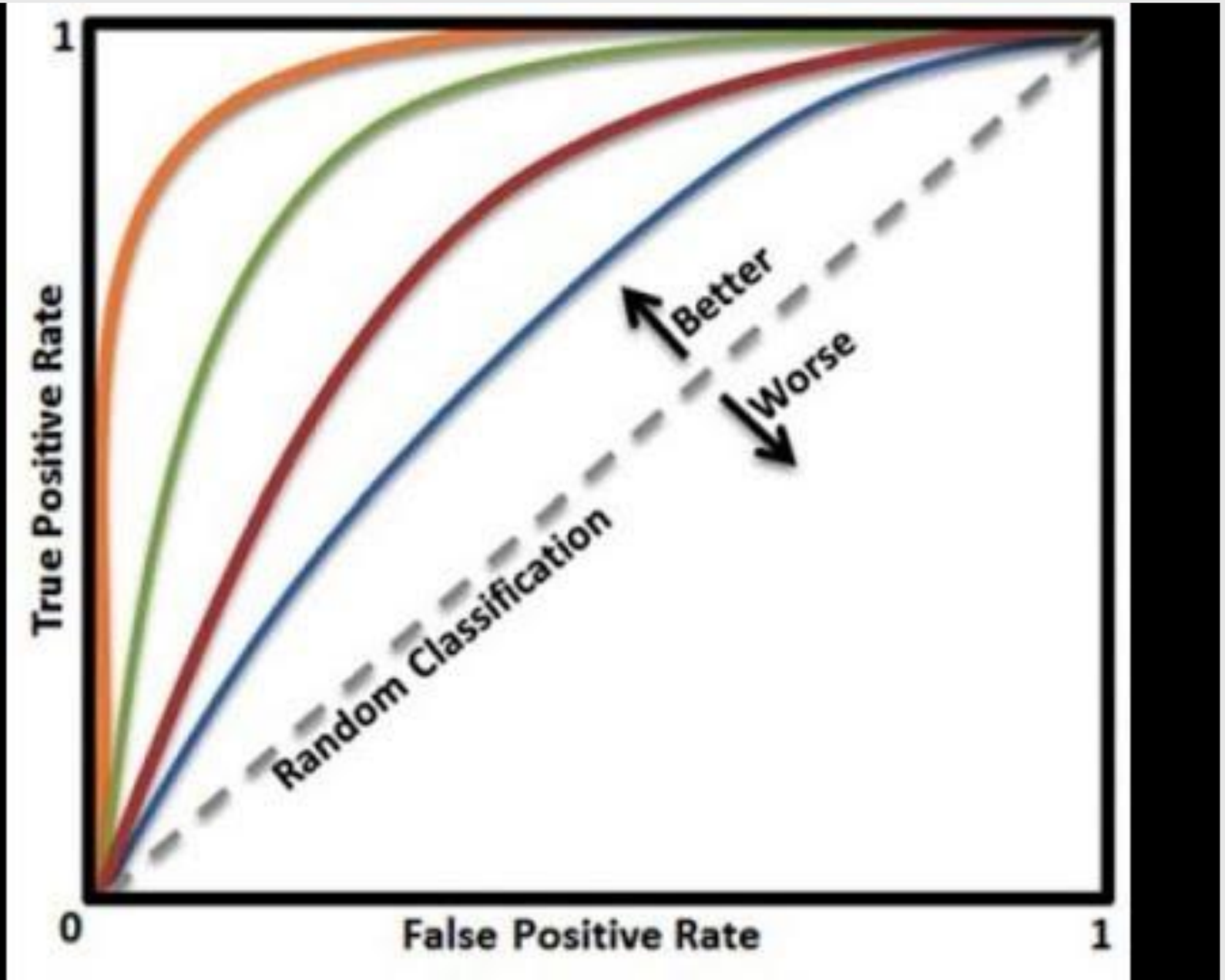
How many relevant items are selected?

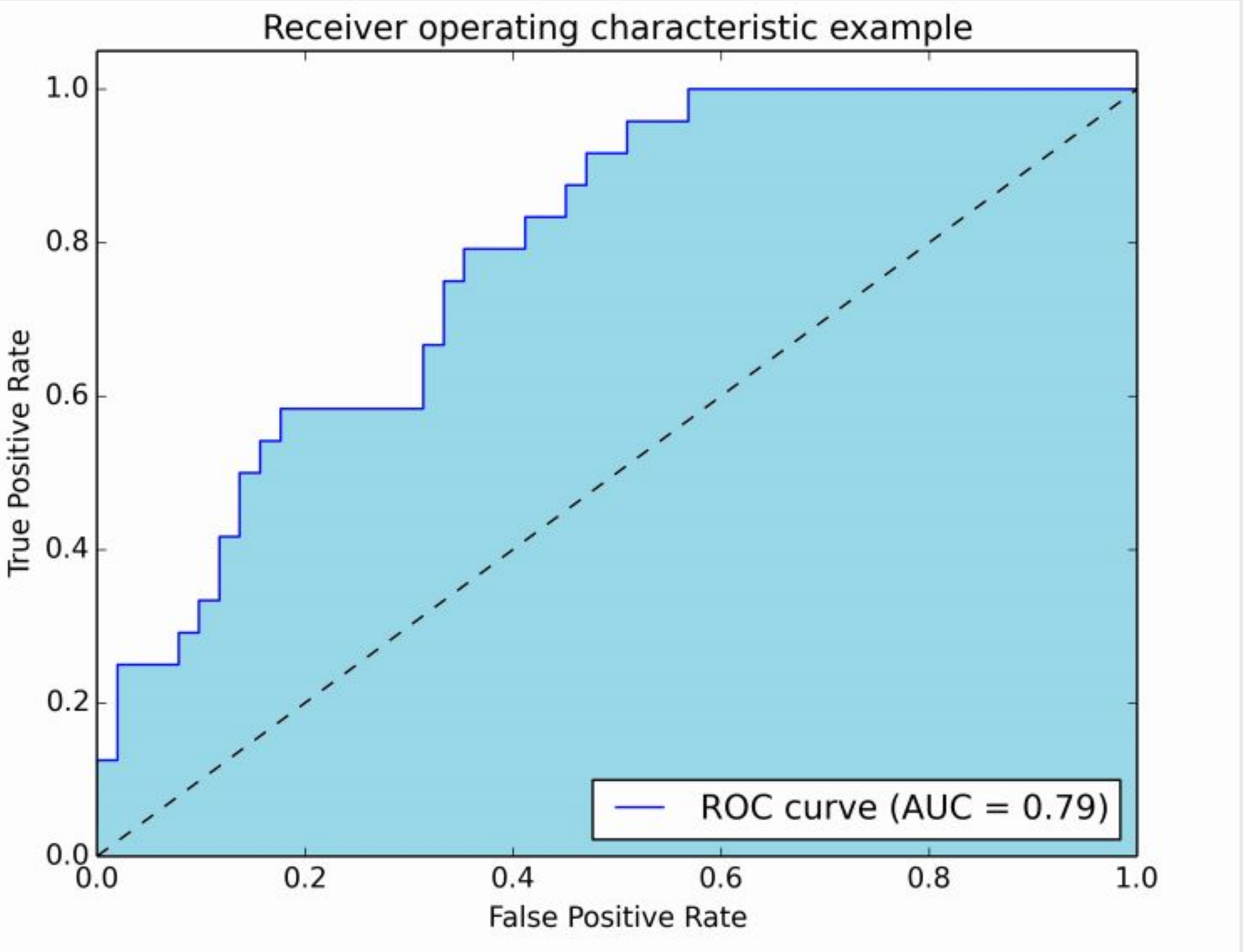
$$\text{Recall} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}}$$


Confusion Matrix
735
180
120
821
False (truth)
True (truth)
False (pred)
True (pred)

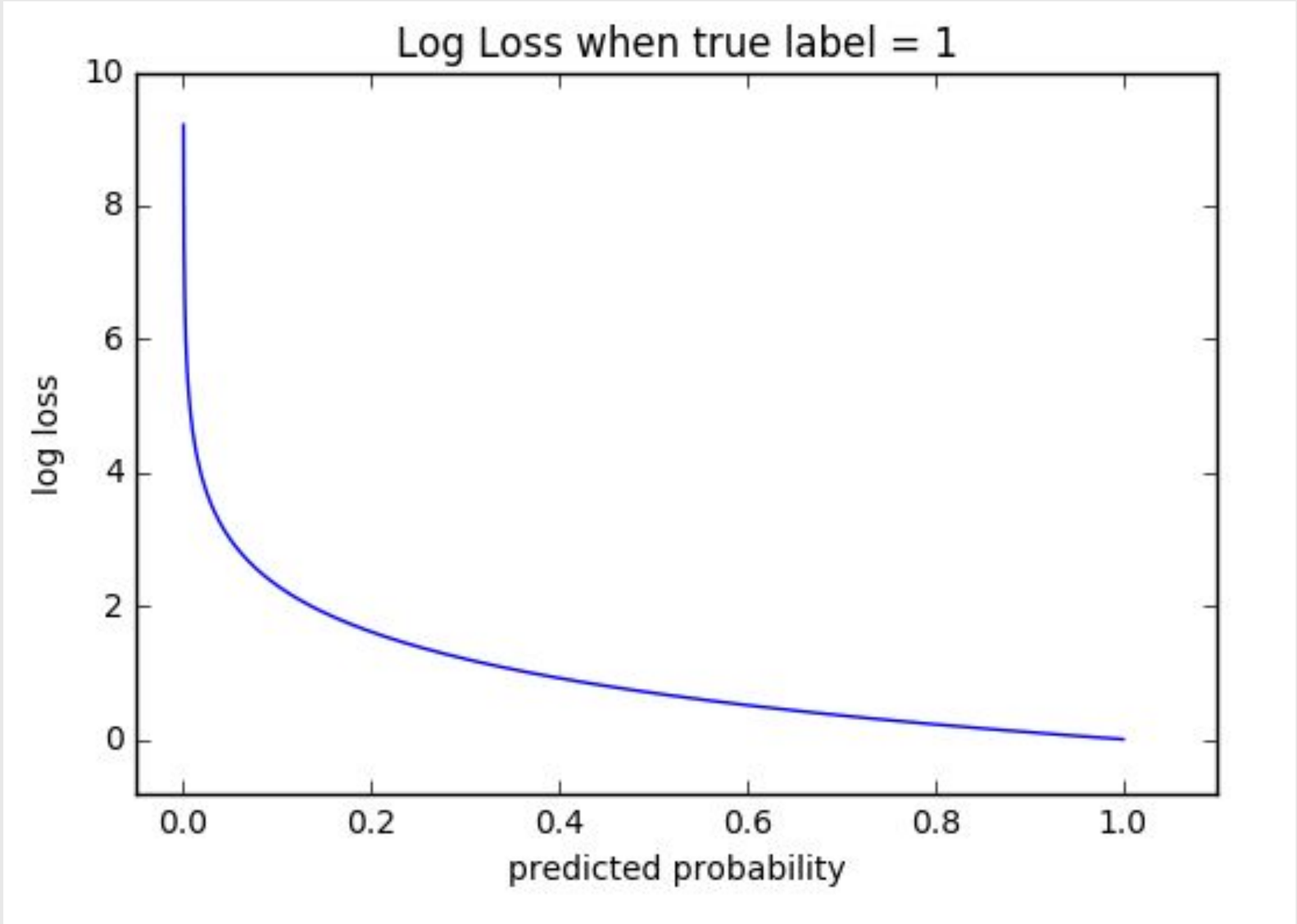
$$F_1 = 2 * \frac{precision * recall}{precision + recall}$$


mcast.
Practice...

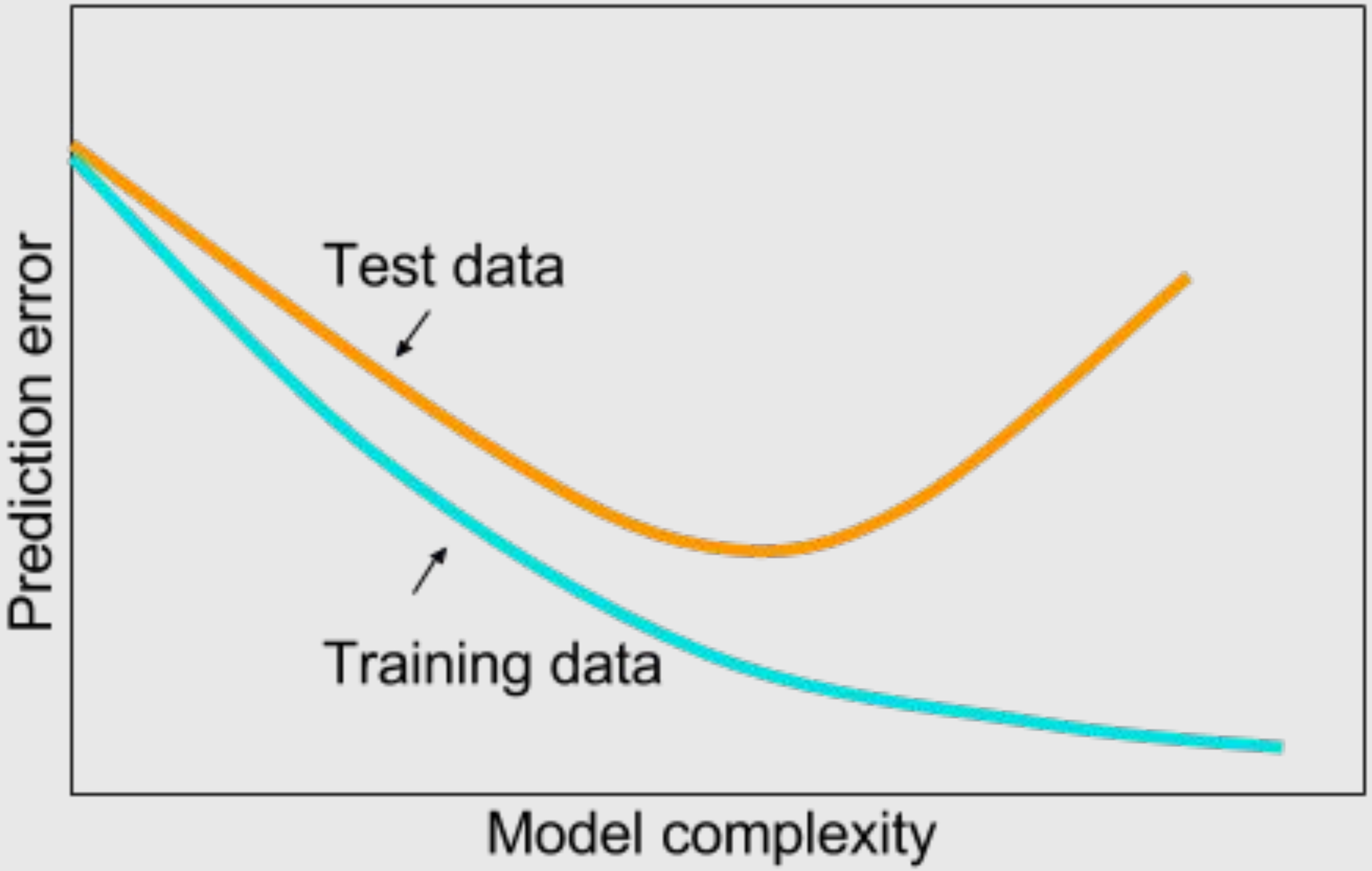

1
True Positive Rate
Better
Worse
Random Classification
0
False Positive Rate
1


Receiver operating characteristic example
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
ROC curve (AUC = 0.79)

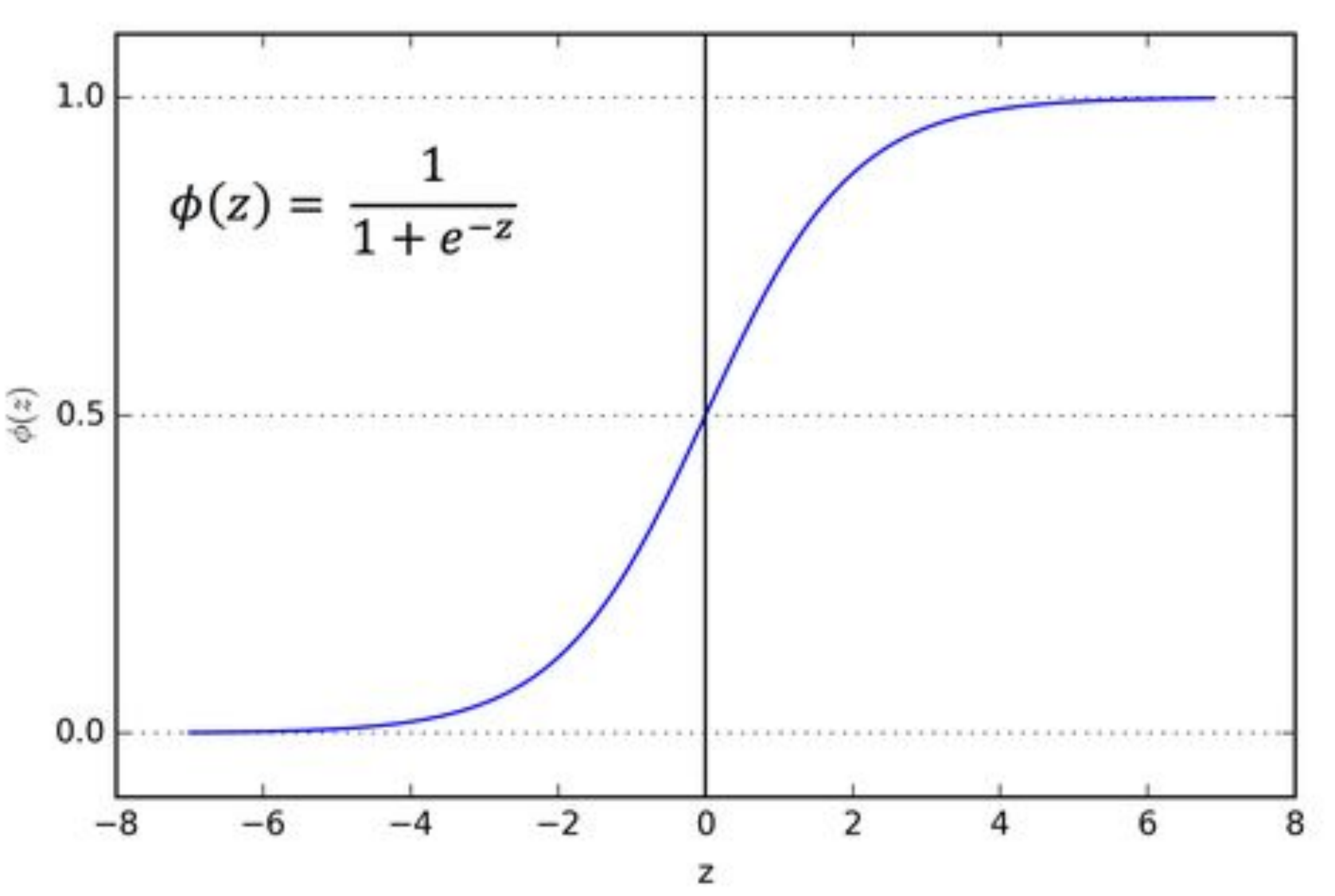
Log Loss


Log Loss when true label = 1
log loss
predicted probability
10
8
6
4
2
0
0.0
0.2
0.4
0.6
0.8
1.0

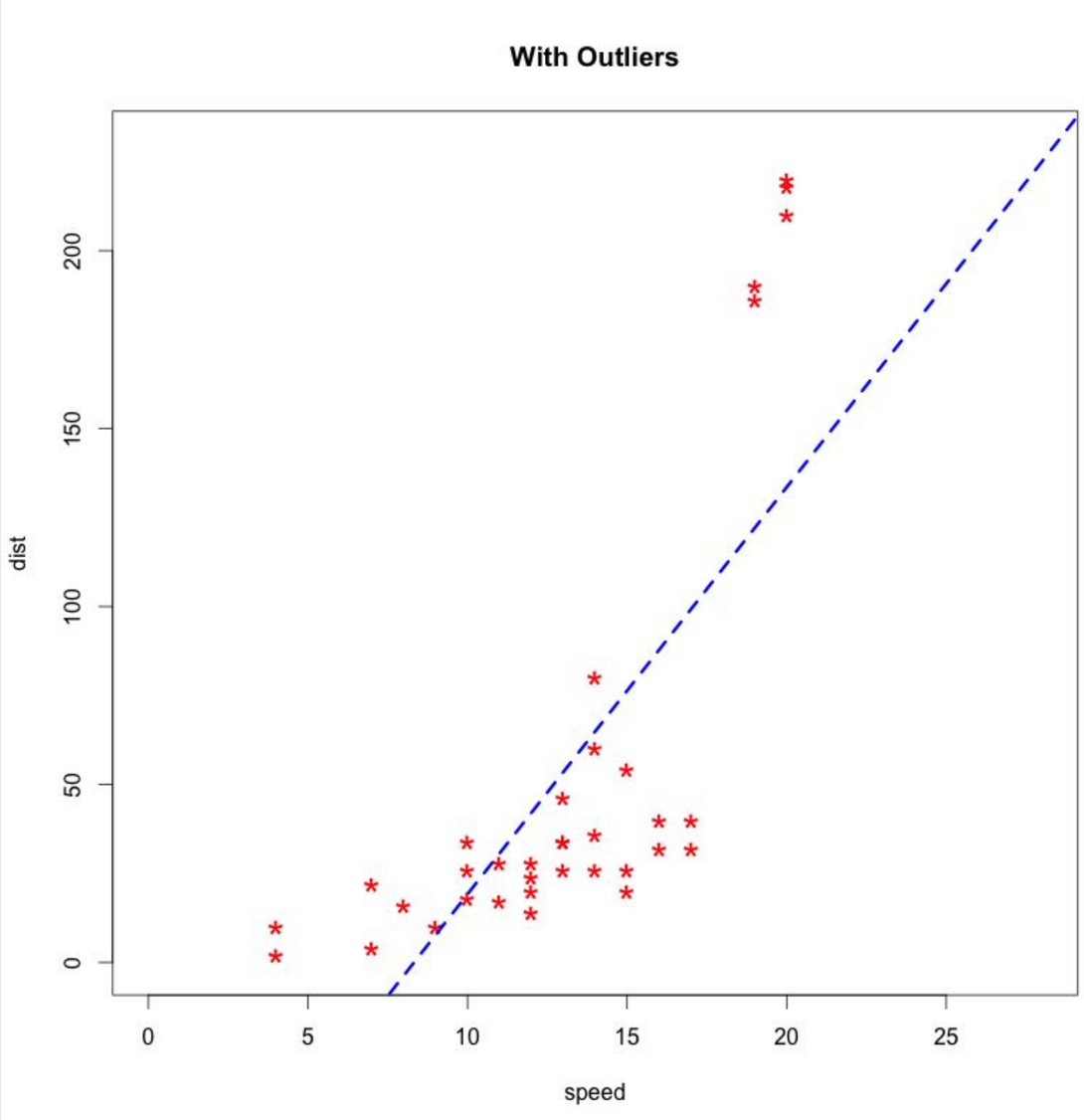
Overfitting

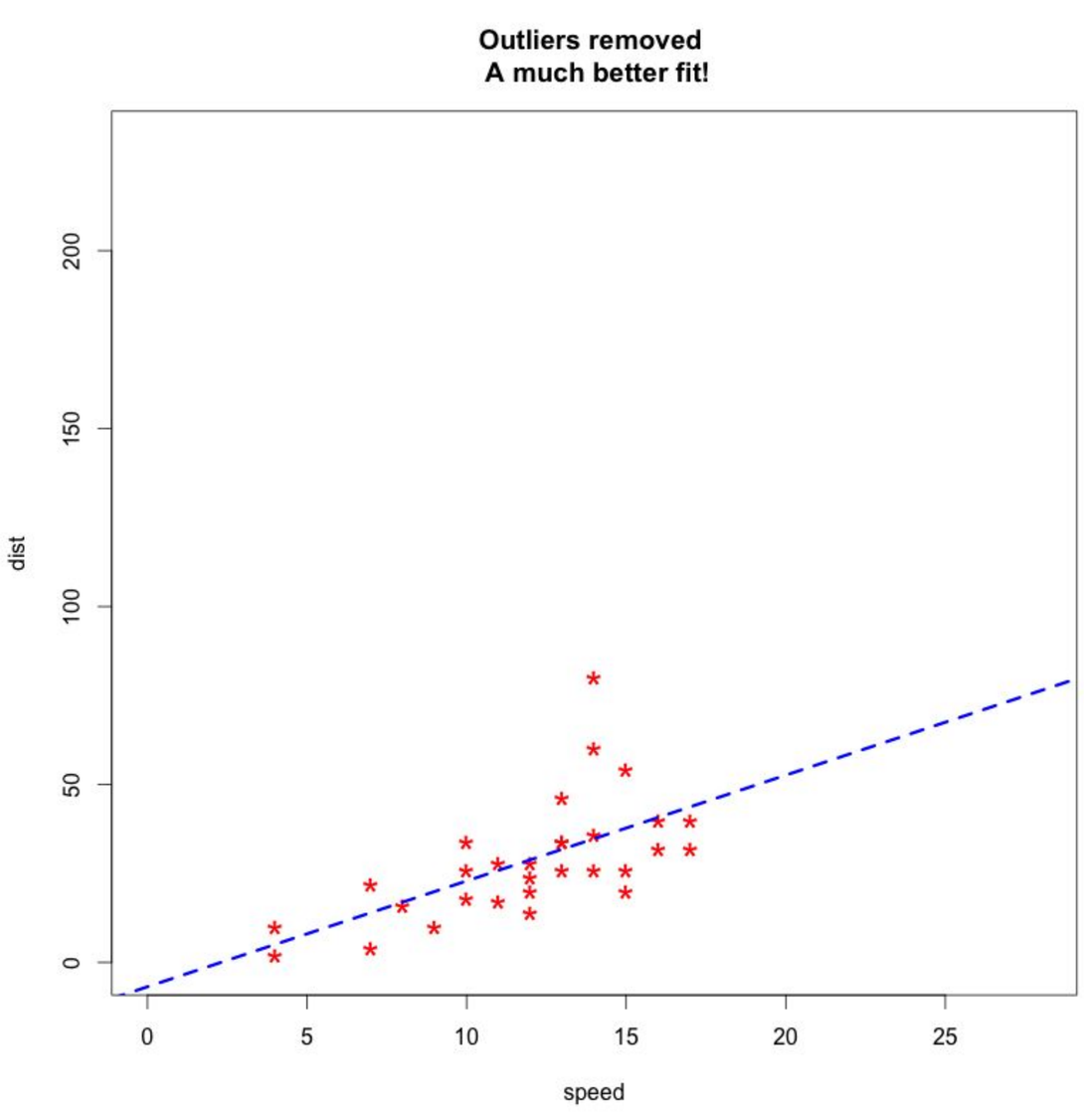


Sigmoid function



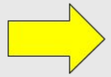
Outliers


With Outliers
dist
speed
0
50
100
150
200
0
5
10
15
20
25


Outliers removed
A much better fit!
dist
speed
0
50
100
150
200
0
5
10
15
20
25

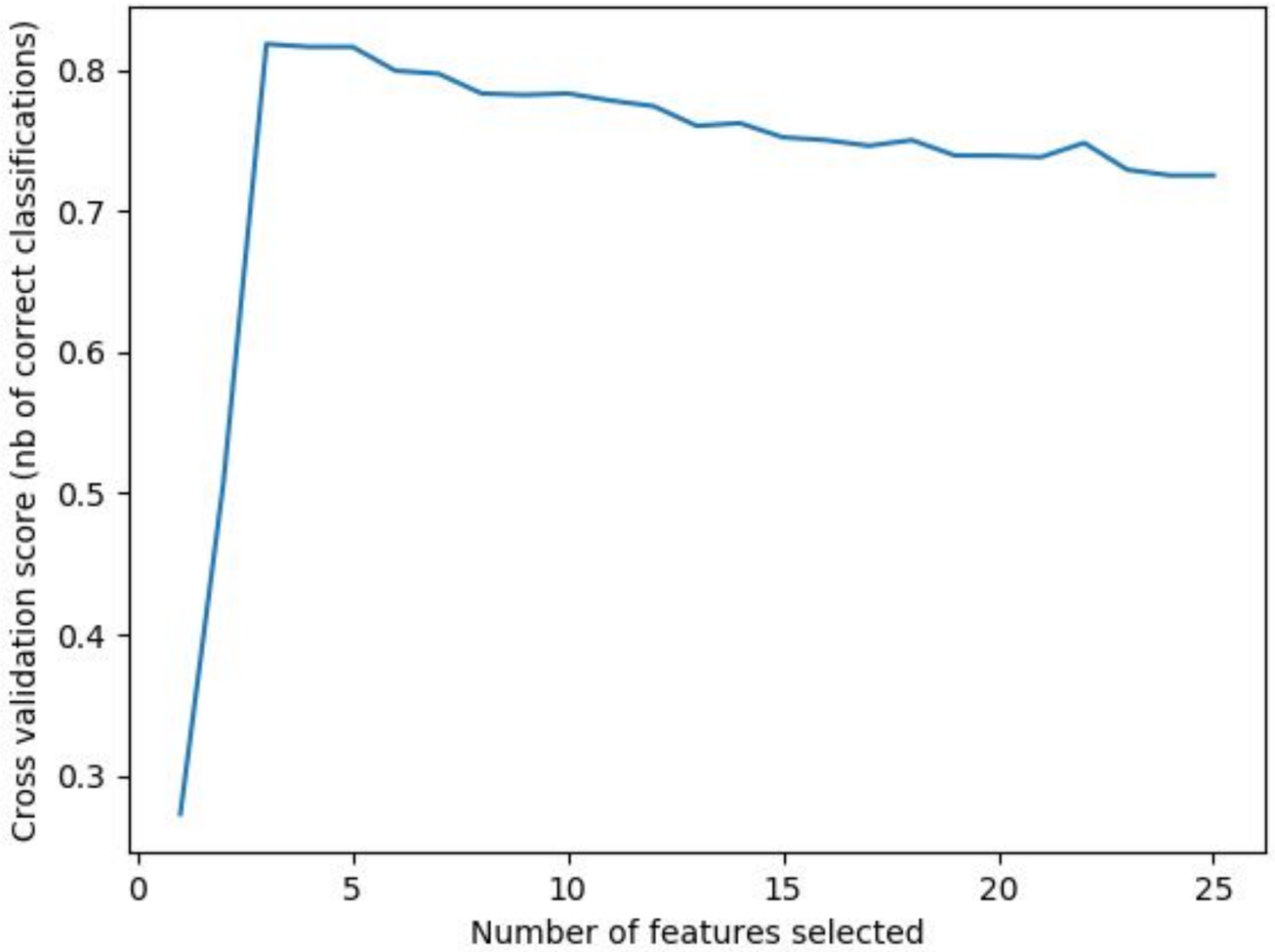
One-hot-encoding

Color
Red
Red
Yellow
Green
Yellow



Red	Yellow	Green
1	0	0
1	0	0
0	1	0
0	0	1

Feature elimination



DÚVIDAS?