Module 4B: Classification Evaluation

Evaluation of Binary Classification

ACCURACY
Proportion of predictions that the model got right

(TN+TP) ÷ (TN+FN+FP+TP)

PRECISION
Proportion of predicted positive cases where the true label is actually positive

TP ÷ (TP+FP)

RECALL
Proportion of positive cases that the model identified correctly

TP ÷ (TP+FN)

F1 SCORE
Overall metric combining Recall and Precision

(2 x Precision x Recall) ÷ (Precision + Recall)

Evaluation of Binary Classification

		Predicted Labels: TRUE	Predicted Labels: FALSE
Real Labels	TRUE	True Positives	False Negatives
Real Labels	FALSE	False Positives	True Negatives

ID	Colour
1	Red
2	Blue
3	Green
4	Green
5	Red

get_dummies (pandas)

ID	Red	Blue	Green
1	1	0	0
2	0	1	0
3	0	0	1
4	0	0	1
5	1	0	0

Converts discrete non-numeric data into binary values

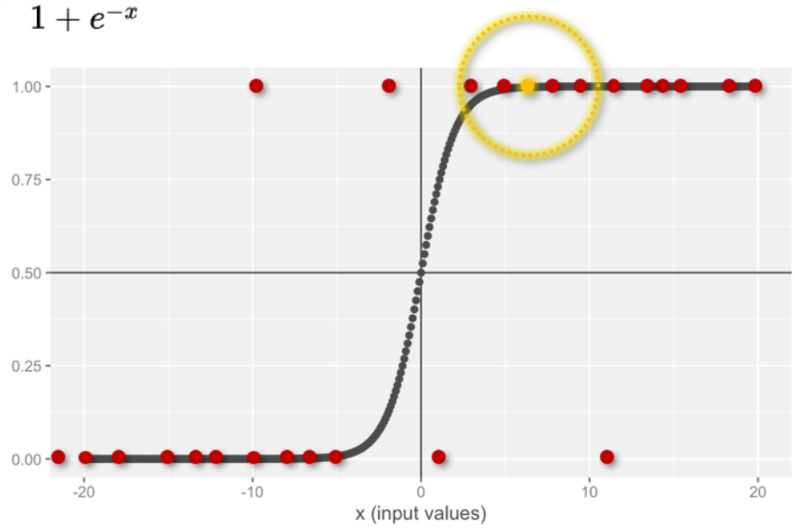
LabelEncoder (scikit-learn)

ID	Colour
1	1
2	2
3	3
4	3
5	1

Converts discrete non-numeric data into discrete numbers

Logistic Function

$$f(x) = \frac{1}{1 + e^{-x}}$$



The logistic function or logistic curve is an S-shaped curve, used in statistics and machine learning to represent logistical regression.

CompSci35_Data Mining
Compiled and created by: Andrei Martin Diamante