taplingua

TALENT COCO

AI/ML Projects/Internships

Hyper Parameter Tuning

Hyper Parameters

- Logistic Regression
  - Learning rate
  - Solvers
- KNN
  - Nearest Neighbors (K)
- Decision Tree
  - Pruning factor (ccp_alpha)

How do you evaluate a model?

Is accuracy alone enough?

Different Metrics

- Accuracy
- Precision
- Recall
- F1 Score

Confusion Matrix

Confusion Matrix

Actual Class	Predicted Class		
		Class = No	Class = Yes
	Class = No		
	Class = Yes		

Deadly virus or not?

Actual Class	Predicted Class		
		Class = No	Class = Yes
	Class = No		
	Class = Yes		

Total = 100
Positive = 50
Negative = 50

Correct = 99
Wrong = 1
Accuracy = 99.9%

Recall

Actual Class	Predicted Class		
		Class = No	**Class = Yes**
	Class = No	True Negative	False Positive
	Class = Yes	False Negative	True Positive

$$\text{Recall} = \frac{TP}{TP + FN}$$

Deadly virus or not?

Actual Class	Predicted Class		
		Class = No	**Class = Yes**
	Class = No	50 TN	0 FP
	Class = Yes	1 FN	49 TP

$$\text{Recall} = \frac{TP}{TP + FN}$$

Correct	=	99
Wrong	=	1
Accuracy	=	99.9%
Recall	=	0.98

Spam Detection

Actual Class	Predicted Class		
		Class = No	Class = Yes
	Class = No		
	Class = Yes		

Total	=	100
Positive	=	50
Negative	=	50

Correct	=	99
Wrong	=	1
Accuracy	=	99.9%
Recall	=	1.00

Precision

Actual Class	Predicted Class		
		Class = No	**Class = Yes**
	Class = No	True Negative	False Positive
	Class = Yes	False Negative	True Positive

$$\text{Precision} = \frac{TP}{TP + FP}$$

Spam Detection

Actual Class	Predicted Class		
		Class = No	Class = Yes
	Class = No		
	Class = Yes		

$$\text{Precision} = \frac{TP}{TP + FP}$$

Correct = 99
Wrong = 1
Accuracy = 99.9%
Recall = 1.00
Precision = 0.98

F1 Score

- Weighted average between precision and recall.
- You can give more weight to precision or recall based on the problem statement

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

Classification Report

********** Test Data Report **********

	precision	recall	f1-score	support
0	0.88	0.83	0.86	218
1	0.73	0.79	0.76	238
2	0.83	0.78	0.80	219
3	0.68	0.69	0.69	225
accuracy			0.77	900
macro avg	0.78	0.77	0.78	900
weighted avg	0.78	0.77	0.78	900

Assignment

- Implement grid search for
  - solvers and C value in your logistic regression model
  - K value for the KNN model
- Create the confusion matrix for the predictions from the decision tree model.
- Display a classification report of your decision tree model
- Justify the performance of the model based on the classification report (Recall, Precision and accuracy)