

7.6 Tutorial- Logistic Regression in Python

Friday, 11 November 2022 16:13

Summary	<ul style="list-style-type: none">Implementing Logistic Regression in PythonUnderstanding Accuracy, Precision and Recall better																										
	<ul style="list-style-type: none">The work is done in the Colab.																										
<ul style="list-style-type: none">Confusion matrix is always _____ matrix.Define the confusion matrix in terms of TN, TP, FN, FP.	<p>Confusion matrix is always a 2x2 matrix.</p> <p>The confusion matrix that we got in the data set:</p> <div><table><tr><th colspan="2" rowspan="2"></th><th colspan="2">Prediction</th></tr><tr><th>0</th><th>1</th></tr><tr><th rowspan="2">Actual</th><th>0</th><td>45323</td><td>3367</td></tr><tr><th>1</th><td>8401</td><td>5335</td></tr></table><table><tr><th colspan="2" rowspan="2"></th><th colspan="2">Prediction</th></tr><tr><th>0</th><th>1</th></tr><tr><th rowspan="2">Actual</th><th>0</th><td>TN</td><td>FP</td></tr><tr><th>1</th><td>FN</td><td>TP</td></tr></table></div> <ul style="list-style-type: none">45323 : The number of times the prediction was 0 and the actual value was also 0: True Negative (TN).5335 : The number of times the prediction was 1 and the actual value was also 1: True Positives (TP).3367 : The number of times the prediction was 1 but the actual value was 0: False Positives (FP).8401 : The number of times the prediction was 0 and the actual value was 1: False Negatives (FN).			Prediction		0	1	Actual	0	45323	3367	1	8401	5335			Prediction		0	1	Actual	0	TN	FP	1	FN	TP
				Prediction																							
		0	1																								
Actual	0	45323	3367																								
	1	8401	5335																								
		Prediction																									
		0	1																								
Actual	0	TN	FP																								
	1	FN	TP																								
<ul style="list-style-type: none">Define: <ol style="list-style-type: none">AccuracyPrecisionRecall	<div><table><tr><th colspan="2" rowspan="2"></th><th colspan="2">Prediction</th></tr><tr><th>0</th><th>1</th></tr><tr><th rowspan="2">Actual</th><th>0</th><td>TN</td><td>FP</td></tr><tr><th>1</th><td>FN</td><td>TP</td></tr></table><div>Accuracy = $\frac{\text{TN} + \text{TP}}{\text{Total number of observations}}$</div><ul style="list-style-type: none">We want to predict the number of 1's in our problems.So, precision is out of all 1 predictions, how many were correct.Precision for predicting 1:<div>Precision = $\frac{\text{TP}}{\text{TP} + \text{FP}}$</div><ul style="list-style-type: none">Recall focuses on actuals.Out of all the actual 1's, how many the model is predicting correctly.Recall for predicting 1:<div>Recall = $\frac{\text{TP}}{\text{TP} + \text{FN}}$</div><ul style="list-style-type: none">Tip to remember Precision and Recall:<div><div></div><div></div><div>Recall</div><div>Precision</div></div></div>			Prediction		0	1	Actual	0	TN	FP	1	FN	TP													
				Prediction																							
		0	1																								
Actual	0	TN	FP																								
	1	FN	TP																								
	<div><table><tr><th colspan="2" rowspan="2"></th><th colspan="2">Prediction</th></tr><tr><th>0</th><th>1</th></tr><tr><th rowspan="2">Actual</th><th>0</th><td>45323</td><td>3367</td></tr><tr><th>1</th><td>8401</td><td>5335</td></tr></table><table><tr><th colspan="2" rowspan="2"></th><th colspan="2">Prediction</th></tr><tr><th>0</th><th>1</th></tr><tr><th rowspan="2">Actual</th><th>0</th><td>TN</td><td>FP</td></tr><tr><th>1</th><td>FN</td><td>TP</td></tr></table></div>			Prediction		0	1	Actual	0	45323	3367	1	8401	5335			Prediction		0	1	Actual	0	TN	FP	1	FN	TP
				Prediction																							
		0	1																								
Actual	0	45323	3367																								
	1	8401	5335																								
		Prediction																									
		0	1																								
Actual	0	TN	FP																								
	1	FN	TP																								

	<ul style="list-style-type: none">• $Accuracy = \frac{TN + FP}{Total\ Number\ of\ observations} = \frac{45323 + 5335}{4523 + 3367 + 8401 + 5335} = \frac{50658}{62426} = 0.81$• $Precision = \frac{TP}{TP + FP} = \frac{5335}{5335 + 3367} = \frac{5335}{8702} = 0.61$• $Recall = \frac{TP}{TP + FN} = \frac{5335}{5335 + 8401} = \frac{5335}{13736} = 0.39$																														
	<pre>print(classification_report(y_test,y_pred))</pre> <table><tr><th></th><th>precision</th><th>recall</th><th>f1-score</th><th>support</th></tr><tr><td>0</td><td>0.84</td><td>0.93</td><td>0.89</td><td>48690</td></tr><tr><td>1</td><td>0.61</td><td>0.39</td><td>0.48</td><td>13736</td></tr><tr><td>accuracy</td><td></td><td></td><td>0.81</td><td>62426</td></tr><tr><td>macro avg</td><td>0.73</td><td>0.66</td><td>0.68</td><td>62426</td></tr><tr><td>weighted avg</td><td>0.79</td><td>0.81</td><td>0.79</td><td>62426</td></tr></table> <ul style="list-style-type: none">• F1-score: weighted average between precision and recall.• Precision and Recall are defined based on the problem that we're trying to address.• Here, we're trying to address 1, so our precision = 0.61, and recall = 0.39.• Whereas, the accuracy is always the same, which is 0.81 here.• Support: number of the data points that we took for the calculations.		precision	recall	f1-score	support	0	0.84	0.93	0.89	48690	1	0.61	0.39	0.48	13736	accuracy			0.81	62426	macro avg	0.73	0.66	0.68	62426	weighted avg	0.79	0.81	0.79	62426
	precision	recall	f1-score	support																											
0	0.84	0.93	0.89	48690																											
1	0.61	0.39	0.48	13736																											
accuracy			0.81	62426																											
macro avg	0.73	0.66	0.68	62426																											
weighted avg	0.79	0.81	0.79	62426																											