

Exercises

§6.8: Classification Performance

Python: There is an optional starter script that supports Python-based problem(s) on this sheet. Use it as scaffolding: read the comments, complete the lines marked **TODO**, and compare the output with the mathematics requested in the problem(s).

Problem 1 A classifier has $TP = 30$, $FP = 10$, $FN = 20$, $TN = 140$. Compute: accuracy, precision, recall, specificity, and F_1 .

Problem 2 A dataset contains 5000 negative observations and 50 positive observations. A classifier predicts every observation as negative. Compute its accuracy and recall. Explain why reporting only accuracy would be misleading.

Problem 3 A medical screening model has $TP = 90$, $FN = 10$, $FP = 180$, $TN = 720$. Compute recall and precision. If the main purpose is to avoid missing disease, which metric is more directly aligned with that purpose?

Problem 4 Suppose a classifier gives

Observation	1	2	3	4	5
y	1	1	0	0	0
p	.90	.45	.70	.35	.10

Compute the confusion matrix at thresholds $t = 0.5$ and $t = 0.3$. For each threshold compute recall and specificity. Describe the tradeoff.

Problem 5 Show algebraically that $FPR = 1 - \text{Specificity}$ and $FNR = 1 - \text{Recall}$.

Problem 6 Suppose a classifier has $\text{Precision} = 0.75$ and $\text{Recall} = 0.60$. Compute F_1 . Compare it with the arithmetic mean of precision and recall.

Problem 7 Use Python to compute TP , FP , FN , TN , accuracy, precision, recall, specificity, and F_1 for

```
y = np.array([1, 0, 1, 1, 0, 0, 1, 0])
p = np.array([.8, .7, .6, .4, .3, .2, .9, .1])
```

using threshold 0.5.

Problem 8 Explain the difference between: (a) training cross-entropy, (b) test-set accuracy, (c) ROC AUC. Why might two classifiers have similar accuracy but different AUC values?