Machine Learning

Kevin Corcoran

December 9, 2021

Contents

Lecture 1: Assessing Classification Performance

Mon 04 Oct 2021 15:01

1 Contingency Tables

	Predicted +	Predicted −	
Actual +	5 (TP)	5 (FP)	10 ($\hat{P}$)
Actual−	4 (FN)	100 (TN)	104 ($\hat{N}$)
	9 (P)	105 (N)	114

1.1 Accuracy

$$\text{acc} = \frac{TP + TN}{P + N} = 92.1\% \tag{1}$$

1.2 Precision

$$\text{prec} = \frac{TP}{\hat{P}} = 50\% \tag{2}$$

1.3 Recall/Sensitivity

$$\text{recall} = \frac{TP}{P} = 55.6\% \tag{3}$$

1.4 Examples

- Covid19
  - Recall is more important
  - Want less false negatives. Don't want someone with covid (pos) to think the don't have it (neg).
- Google
  - Precision is more important
  - Want less false positives. Don't want an unwanted (pos) to show up in the search results.

1.5 Contingency Venn Diagram

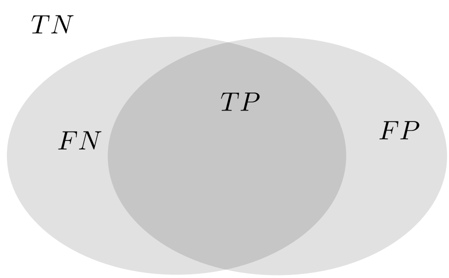

Figure 1: Contingency Venn Diagram

2 Coverage Plots

Consider classifiers C_1, C_2, C_3

In a coverage plot, classifiers with the same accuracy are connected by line segments with slope 1. Both C_1 and C_3 dominate C_2, but the choice of classifier depends on whether we want more true positives (less false negatives), in which case we choose C_3, or we care less about true positives and want less false positives, in which case we choose C_1.

So for Covid19, we choose C_3, and for Google, we choose C_1?

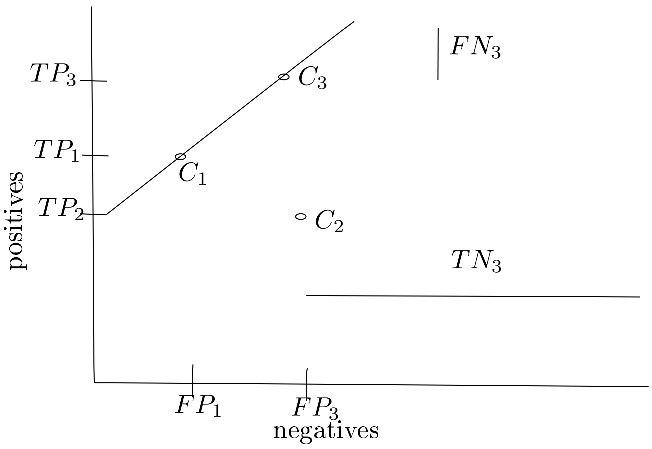

Figure 2: Coverage Plot