

28 June 2021 Report

Brad Burkman

26 June 2021

Contents

1	Accomplishments This Week	1
2	Accuracy, Precision, Recall, Sensitivity, and f1 with Imbalanced Data	1
2.1	Reminder of Definitions of Metrics	1
2.2	Imbalanced Data Set	2
2.3	The Problem	2
2.4	Balanced Accuracy	2
2.5	Balanced Precision	3
3	References	4

1 Accomplishments This Week

-

2 Accuracy, Precision, Recall, Sensitivity, and f1 with Imbalanced Data

2.1 Reminder of Definitions of Metrics

		Prediction	
		N	P
Actual	N	TN	FP
	P	FN	TP

$$\text{Accuracy} = \frac{TN + TP}{TN + FP + FN + TP}$$

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

$$\text{Specificity, Selectivity, or TNR} = \frac{TN}{TN + FP}$$

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

2.2 Imbalanced Data Set

In an unbalanced data set, the number of actual negatives ($N = TN + FP$) is much different from the number of actual positives ($P = FN + TP$). In our case, if our independent variable is fatal crashes, the negatives are 99.574714% of the data set, and the positives are just 0.425286%.

2.3 The Problem

The standard metrics get thrown off by the imbalance. If we predict that every crash is nonfatal, we have accuracy of 99.57%, which sounds really impressive.

The recall (true positive rate) is not thrown off by an imbalanced data set, because it only works with TP and FN, the actual positives. Similarly for specificity (true negative rate).

The precision is thrown off by an imbalanced data set, because it works with both a subset of the actual positives (TP) and a subset of the actual negatives (FP).

2.4 Balanced Accuracy

There is a metric called *balanced accuracy*. You get it from the definition of *accuracy* by multiplying the actual negative elements (TN and FP) by the ratio of the positives to negatives,

$$\frac{P}{N} = \frac{FN + TP}{TN + FP}$$

so that the total number of actual negatives and total number of actual positives in the sample are equal.

[I suppose you could also get it by multiplying the actual positive elements (FN and TP) by the reciprocal.]

I got this derivation by intuiting about what I would want *balanced accuracy* to mean, and it matches the definition I found in Wikipedia.

https://en.wikipedia.org/wiki/precision_and_recall#Imbalanced_data

Wikipedia says [I'm sure I can find a more authoritative source.]

$$\text{Balanced Accuracy} = \frac{TPR + TNR}{2}$$

$$\begin{aligned}
\text{Recall or TPR} &= \frac{TP}{TP + FN} \\
\text{Specificity or TNR} &= \frac{TN}{TN + FP} \\
\text{Accuracy} &= \frac{TN + TP}{TN + FP + FN + TP} \\
\text{Balanced Accuracy} &= \frac{TN \cdot \frac{P}{N} + TP}{TN \cdot \frac{P}{N} + FP \cdot \frac{P}{N} + FN + TP} \\
&= \frac{TN \cdot P + TP \cdot N}{TN \cdot P + FP \cdot P + FN \cdot N + TP \cdot N} \\
&= \frac{TN \cdot P + TP \cdot N}{(TN + FP) \cdot P + (FN + TP) \cdot N} \\
&= \frac{TN(FN + TP) + TP(TN + FP)}{(TN + FP)(FN + TP) + (FN + TP)(TN + FP)} \\
&= \frac{TN(FN + TP) + TP(TN + FP)}{2(TN + FP)(FN + TP)} \\
&= \frac{TN(FN + TP)}{2(TN + FP)(FN + TP)} + \frac{TP(TN + FP)}{2(TN + FP)(FN + TP)} \\
&= \frac{TN}{2(TN + FP)} + \frac{TP}{2(FN + TP)} \\
&= \frac{TNR + TPR}{2}
\end{aligned}$$

2.5 Balanced Precision

I haven't found *balanced precision* in a brief Google search, although Google knows the kind of stuff I look up and sent me to articles on balanced accuracy. Finding it will take some work, because "balanced precision" has different meanings in other tech fields.

We can make balanced precision the same way we made balanced accuracy, by taking the actual negative results (TN and FP) and scaling them so that the total number of actual negatives equals the total number of actual positives, by multiplying by $\frac{P}{N} = \frac{FN+TP}{TN+FP}$.

Is this related to the G-mean in last week's report?

$$\text{G-mean} = \sqrt{\text{Precision} \times \text{Specificity}}$$

$$\begin{aligned}
\text{Precision} &= \frac{TP}{TP + FP} \\
\text{Balanced Precision} &= \frac{TP}{TP + FP \cdot \frac{P}{N}} \\
&= \frac{TP \cdot N}{TP \cdot N + FP \cdot P} \\
&= \frac{TP(TN + FP)}{TP(TN + FP) + FP(FN + TP)}
\end{aligned}$$

3 References