Metrics for one-class classification

Asked 8 years, 1 month ago Modified 6 years, 9 months ago Viewed 8k times

9

How do you calculate precision and recall in one class classification? In other words in one class classification, we just have TP(True Positive) and FN(False Negative). Which metrics we should use for these type of classification?

classification outliers anomaly-detection precision one-class

Share Cite Edit Follow Flag

edited Dec 22, 2016 at 9:42
Tim ♦ 138k 26 257 498

asked Jan 26, 2016 at 13:26
Sara Nikdel 91 1 3

During the training or the testing? – Leila Apr 15, 2019 at 18:38

4 Answers

Sorted by: Highest score (default)

3

We use one-class classification is used when we have only "positive" labels (although some argue for using it when the quality of the data about the labels is poor) for outlier, or anomaly, detection.

With such data you *cannot* assess accuracy of the predictions. Technically you can check if it properly labeled all your data as "positive", but then you would conclude that the useless model that *always* returns "positive" label no matter of data, has perfect fit.

To judge performance of such classifier you would need to have data with "negative" labels. One thing you could do is to simulate data with artificially introduced anomalies (this is often done, e.g. in image classification where you add noise to the data, or transform the images), or simulate such data that you *know* that should be classified as anomaly, and use such data for testing.

The story is different if you *have* data about "positive" and "negative" classes, since then you can use exactly the same tools for evaluating your model as for classification in general, but then, why would you use one-class classification algorithms?

Share Cite Edit Follow Flag

edited Apr 13, 2017 at 12:44
Community Bot 1

answered Dec 22, 2016 at 9:41
Tim ♦ 138k 26 257 498

3

Though it's a late reply, I'd like to point of implicit assumptions by previous answers that likely don't hold.

1. for one-class classification, we don't know the real ratio of positive and negative data. So we cannot any development set has similar distribution to the real data.
2. A standard setting for one-class classification is we have positive and unlabeled dataset. We can't assume we have the label for "negative" data even in the development set. Also, we can't assume all the unlabelled data are "negative".

An alternative evaluation is proposed in the following paper (section 4):

Lee, Wee Sun, and Bing Liu. "Learning with positive and unlabelled examples using weighted logistic regression." ICML. Vol. 3. 2003.

They uses

$$\frac{r^2}{\Pr(Y=1)}$$

P.S.: Prof. Lee, Prof Liu and Dr. Cheng are the people that coined one-class classification. We can take their evaluation as somewhat "official".

Share Cite Edit Follow Flag

edited May 13, 2017 at 16:16
Tim ♦
138k 26 257 498

answered Dec 22, 2016 at 5:42
yiping
76 4

1 Could you explain what r^2 is in here? – Tim ♦ May 13, 2017 at 16:20

Sorry for the late reply. *r* is just the recall. More specifically, *r = Pr[f(x)=1|y=1]*. This is based on section 4 of the reference paper above. – yiping May 29, 2017 at 3:33

I think the formula is $r^2/Pr(f(X)=1)$. – Hossein May 16, 2022 at 15:14

0

@user3791422 has the right answer. In addition, I would like to point out:

If you have the notion of True Positive and False Negative, it means you have a notion of the ground truth and you have predicted responses. Therefore, by definition, False Positive and True Negative should exist.

Logically, what the OP didn't consider is, if we know some examples belong in the class we know the remaining examples (rest of the world population for which we don't actually need training examples) belong outside the class. Therefore, FP and TN can be calculated when we observe them during testing.

Share Cite Edit Follow Flag

edited Mar 6, 2016 at 19:31

answered Mar 3, 2016 at 21:36
Rahul Murmuria
111 3

-1

When you do one class classification, besides TP and FN, you should also have FP (false positive) and TN (true negative). FP are the instances that you have classified as positives when they were actually negative and TN are the instances that you have correctly classified as negatives. Then you can calculate precision and recall:

precision = TP/(TP + FP) recall = TP/(TP + FN)

The wikipedia page https://en.wikipedia.org/wiki/Precision_and_recall explains it very well.

However, when you do one class classification, some other common metrics are the false positive rate (FPR) and the f1-score.

FPR = FP/(FP+TN)

F1-SCORE = 2TP/(2TP+FP+FN)

I hope that helped. Regards!

Share Cite Edit Follow Flag

edited Jan 27, 2016 at 8:30

answered Jan 27, 2016 at 8:23
user3791422
59 1 5

6 In one-class classification you typically don't have negatives, hence its name. What you describe is standard supervised learning. – Marc Claesen Jan 27, 2016 at 8:38

1 As far as I understand it, one-class classification means that you classify the new records according to they are part of the class X or not. So, if we consider a positive as not being from the class, a negative would be being from the class. Supervised vs unsupervised is a different thing than one-class or not. This has to do with having labelled data in the training phase or not. For the one-class algorithm you do not normally have labelled data to train. But, when you want to calculate some metrics I believe that you need some ground truth to be able to calculate the metric. – user3791422 Jan 27, 2016 at 9:56