Evaluation of a Classifier

Accuracy = Proportion of examples with correct prediction

\[ Acc = \frac{\text{Count}(y_d = \hat{y}_d)}{\text{Num docs in test set}} \]

\[ = \sum_{d \in \text{Test set}} 1 \epsilon y_d = \hat{y}_d \]

\[ \frac{1}{|\text{Test Set}|} \]

Train

Dev/Tuning

Test

Fix \( \alpha \)

Make

or: \( X-Val \)

Predict

Acc eval metric

UVMU 3355 QVAV / L
Overusing test set $\Rightarrow$ subtle overfitting.

Does test set acc generalize?

$\alpha = 1e6$
Accuracy = \frac{4}{7}

Confusion Matrix

\begin{array}{cc|c}
\text{Gold} & + & - \\
\hline
\text{Pred} & + & 2 & 2 \\
& - & 1 & 2 \\
\end{array}

Precision & Recall

\text{Precision} = \frac{TP}{TP + FP} = \frac{2}{4} = 0.5

\text{Recall} = \frac{TP}{TP + FN} = \frac{2}{5}
Sarcasm Detection — UMass CS 585 in-class exercise, 9/19/17

These are Reddit comments (from https://arxiv.org/pdf/1704.05579.pdf). Their authors tagged them as “sarcastic” or “not sarcastic”. These tags are hidden. The task is to predict whether the message was tagged as sarcasm. Label them 1=sarcastic, 0=not sarcastic.

1. but CNN told me the leaks are all faked by Russia!

2. Is this some sort of mug shot mash up?

3. there isn't... the website says specifically there is no way to change it once placed, you must cancel the order and place a new one to make any modifications.

4. All men are handsome!

5. Yeah but he paid his dues so it's his turn.

6. ........what?

7. Maybe you should stop reading it

8. It's okay, it's just his opinion!

9. I know that's what she's doing now, I'm saying this isn't the first time I've seen her.

10. honestly i'd have turned around and sold it to buy the K. But at the end of the day, if you aren't really one to overclock, then what difference does it make to you anyway?