

When better  
is worse

Ian Calvert (@iancal)

Example



?



Lemon

Test!



?



Norwich Terrier



?





Cardigan Corgi



?



Leopard

How did  
you do?

Accuracy

$$\frac{TP+TN}{TOTAL}$$

Precision

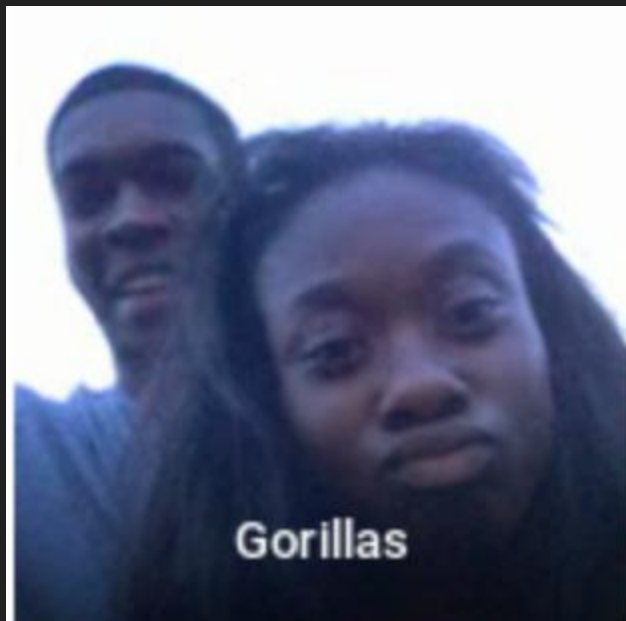
$$\frac{TP}{TP+FP}$$

Recall

$$\frac{TP}{TP+FN}$$

Errors are  
different





Google images 2015

What can we  
do?

Four words

What if we're  
wrong?

# References

Sofa issue:

<http://rocknrollnerd.github.io/ml/2015/05/27/leopard-sofa.html>

Test image data:

<https://cs.stanford.edu/people/karpathy/ilstvrc/>

Google racism tweet:

<https://twitter.com/jackyalcine/status/615329515909156865/photo/1>