1 | Validation

We have visualized out models and used human judgment to, well, judge them. We have *not* done this algorithmically or mathematically.

1.1 | Why?

Some things are blind to the human eye. Eg. underfitting and overfitting.

Not enough data, the algorithm was buggy (can't we see these though? maybe just not as easily?)

1.1.1 | Underfitting

Wrong algorithm, buggy, or the data just sucks / there isn't actually a correlation.

1.1.2 | Overfitting

Training *to well* to our dateset, making it not applicable to the real world / other data.

1.2 | Bias-Variance Tradeoff

Bias - off Variance - inconsistent

We want low bias low variance (doih).

1.3 | Holdout? nah, let's cross validate!

Like holdout, but you do it multiple times with different chunks of data 'held out'

1.4 | Validation?

What do? - Accuracy - Easy, but not super effective / informative. - Precision, Recall, F-measure - True positive, false negative, and all the permutations. - Precision = $\frac{TP}{TP+FP}$ - Recall