Source:

---

1 | **Validation**

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

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?)

Underfitting

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

Overfitting

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

Bias-Variance Tradeoff

Bias - off Variance - inconsistent

We want low bias low variance (doih).

Holdout? nah, let's cross validate!

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

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

```
plant_a_heights = numpy.random.normal(loc=PLANT_A_AVG_HEIGHT, size=NUM_INPUTS)
plant_a_widths = numpy.random.normal(loc=PLANT_A_AVG_WIDTH, size=NUM_INPUTS)

plant_b_heights = numpy.random.normal(loc=PLANT_B_AVG_HEIGHT, size=NUM_INPUTS)
plant_b_widths = numpy.random.normal(loc=PLANT_B_AVG_WIDTH, size=NUM_INPUTS)

plant_c_heights = numpy.random.normal(loc=PLANT_C_AVG_HEIGHT, size=NUM_INPUTS)
plant_c_widths = numpy.random.normal(loc=PLANT_C_AVG_WIDTH, size=NUM_INPUTS)
```

```
# this creates a 2-dimensional matrix, with heights in the first column and widths in the second
# the first half of rows are all plants of type a and the second half are type b
plant_inputs = list(zip(numpy.append(plant_a_heights, plant_b_heights),
                        numpy.append(plant_a_widths, plant_b_widths),
                        numpy.append(plant_a_heights, plant_b_heights)))

# this is a list where the first half are 0s (representing plants of type a) and the second half are 1s
classes =  +  +

‘‘‘
```