UNIVERSITY OF
SAN FRANCISCO

Master of Science
in Analytics

Introduction

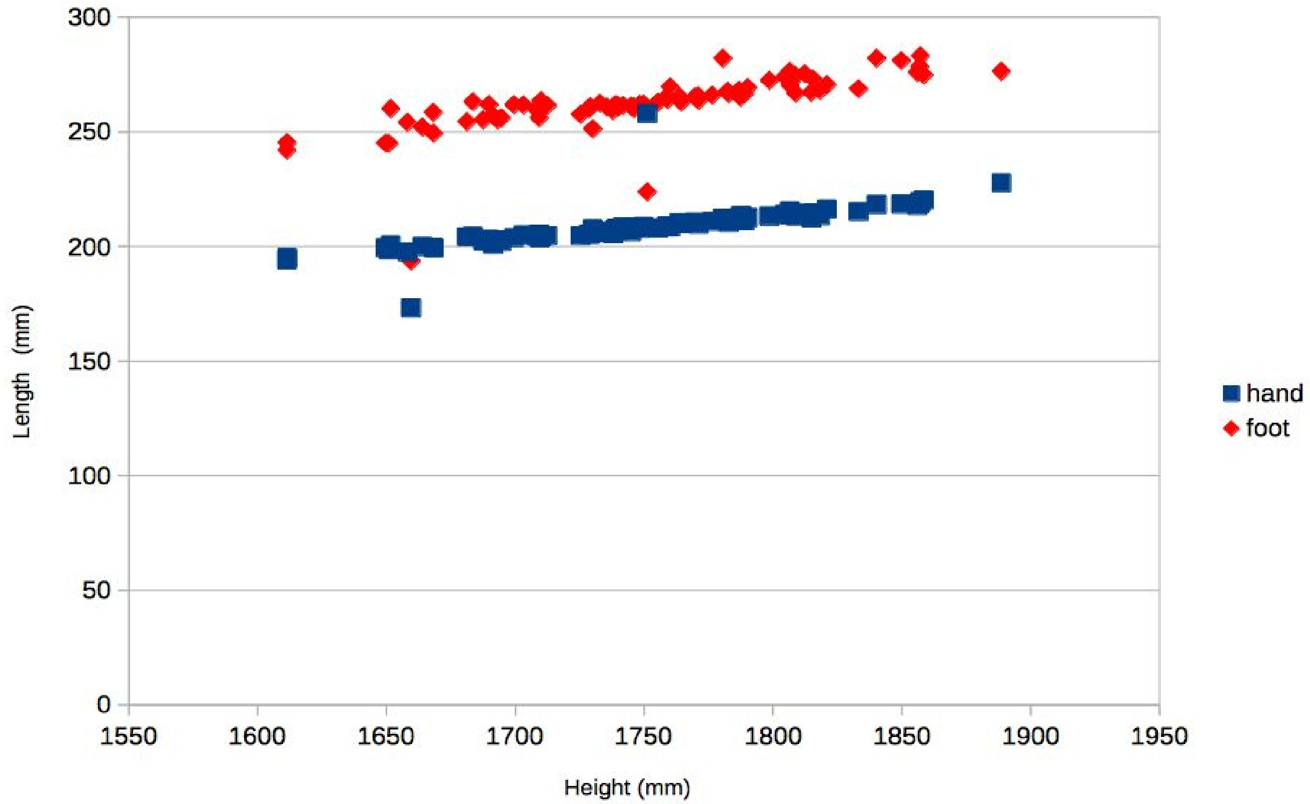
Machine Learning 1

What you will need to know…

- Supervised Learning
    - Regression
        - Quantitative (target is a number)
        - Example: how much will my apartment cost in 5 years?
    - Classification
        - Qualitative (target is a category)
        - Example: is this email spam?
- Unsupervised Learning
    - Principal Components Analysis*
    - Clustering
        - Example: Do these things belong to the same category?
- Other topics (time permitting)
    - Neural Networks
    - Deep Learning

Example (regression)

- Data for 80 males -- height vs. length of hand | foot:



- Source: http://www.stat.ufl.edu/~winner/datasets.html

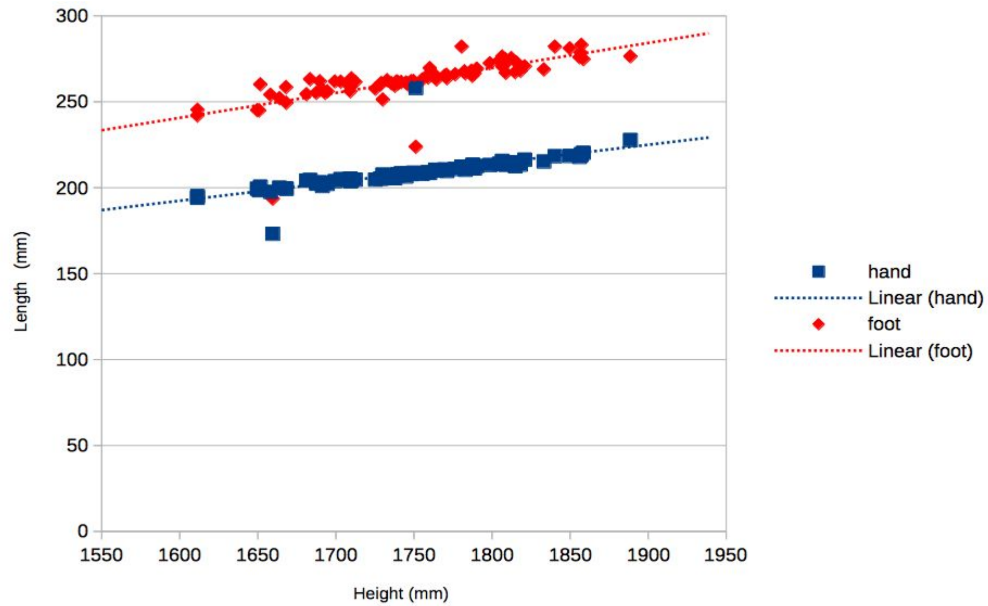
What you will be able to do...

- Prediction
    - Given already seen data, predict some unseen event
- Inference
    - Given already seen data, explain what led to a particular outcome
    - For example: what can you measure to determine an outcome?
    - Or for example: what is the relationship between outcomes and what's measurable?

Basic (regression) concept & terminology

- Given data ($n \times p$), learn best functions & parameters
- Form of function:

$$Y = f(x) + \epsilon$$



- Learn function:

$$\hat{Y} = \hat{f}(x)$$

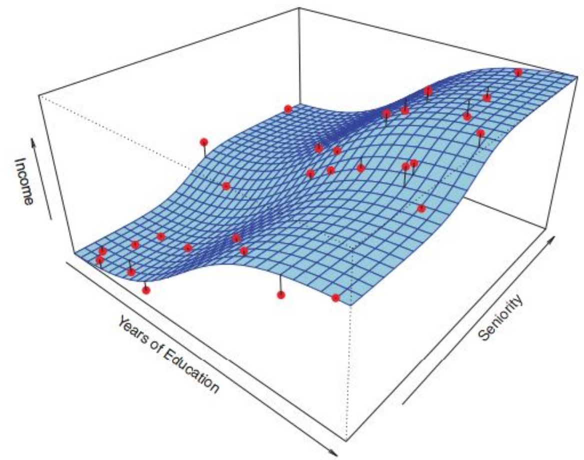
- Y = outcome | { *dependent* | *response* | *output* } variable
- X = predictor(s) | independent variable(s) | feature(s)
- ϵ = error term

Approaches for estimating *f*

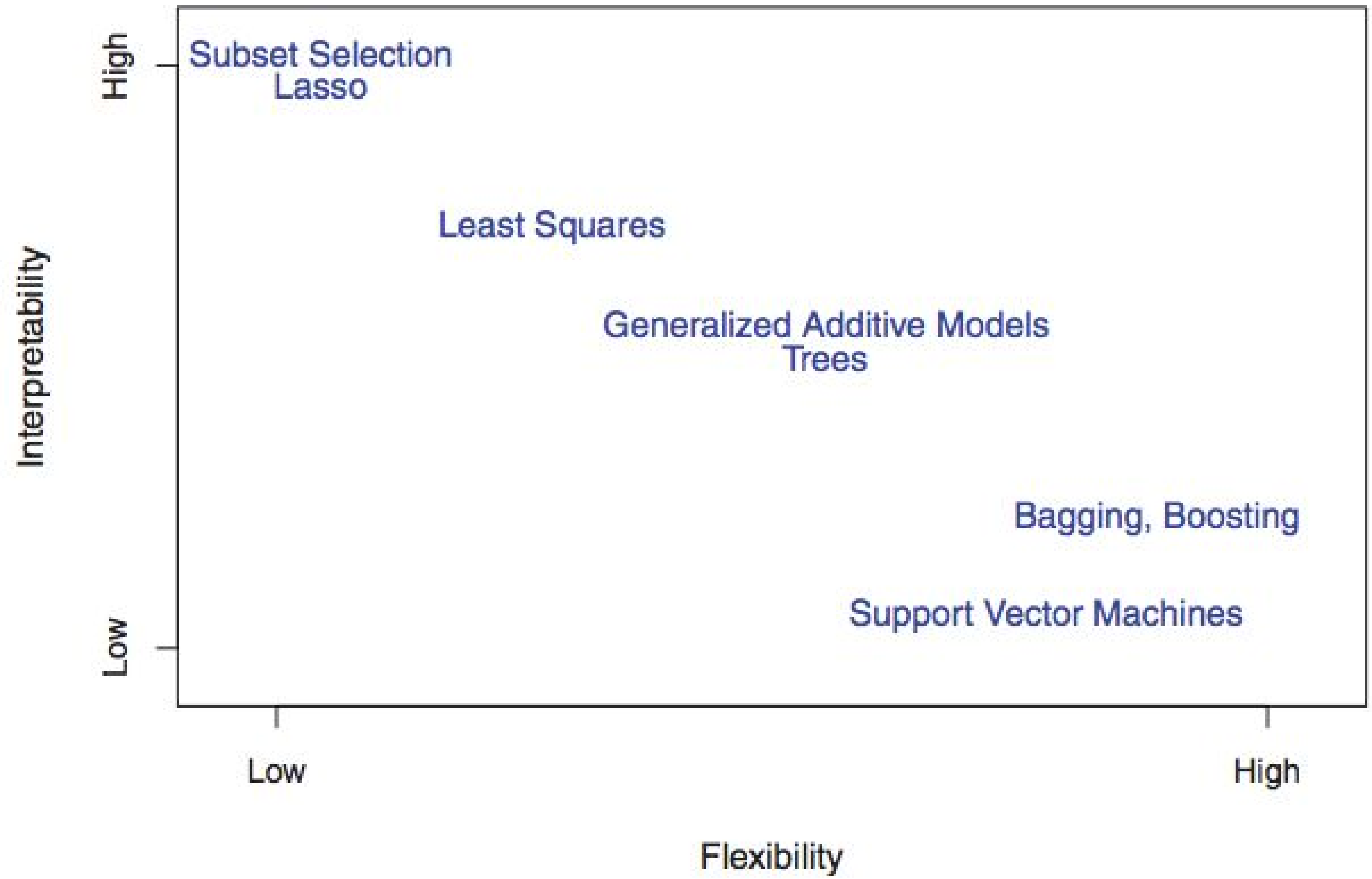
- Parametric
  - Assumption: X and Y have a linear relationship, ala

$$Y \approx \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p$$

  - Estimate parameters β_i
- Non-parametric
  - Generates a (thin / rough) *spline* with no assumptions about relationship between X and Y



Flexibility vs. interpretability



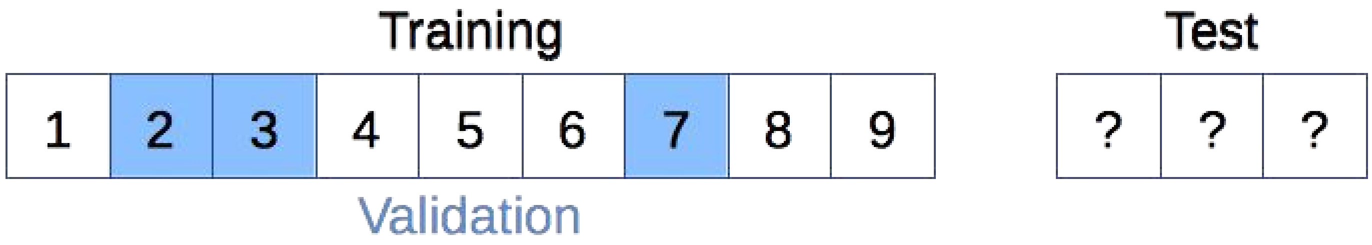
Errors

- Always present (a/k/a “noise”)
- Types
    - Reducible - Can be improved by choosing better *f(x)* / parameters
    - Irreducible - Cannot be improved (due to natural variation in data?)
- Can be quantified?

$$
\begin{aligned}
E(Y-\hat{Y})^2 &= E[f(X)+\epsilon-\hat{f}(X)]^2 \\
&= \underbrace{[f(X)-\hat{f}(X)]^2}_{\text{Reducible}} + \underbrace{\mathrm{Var}(\epsilon)}_{\text{Irreducible}}
\end{aligned}
$$

- Overfitting
    - Function follows data (including errors) too closely

Set aside data for estimating performance



Popular metrics for quantifying model efficacy

- Regression → MSE
  - Mean squared error (on the test set)

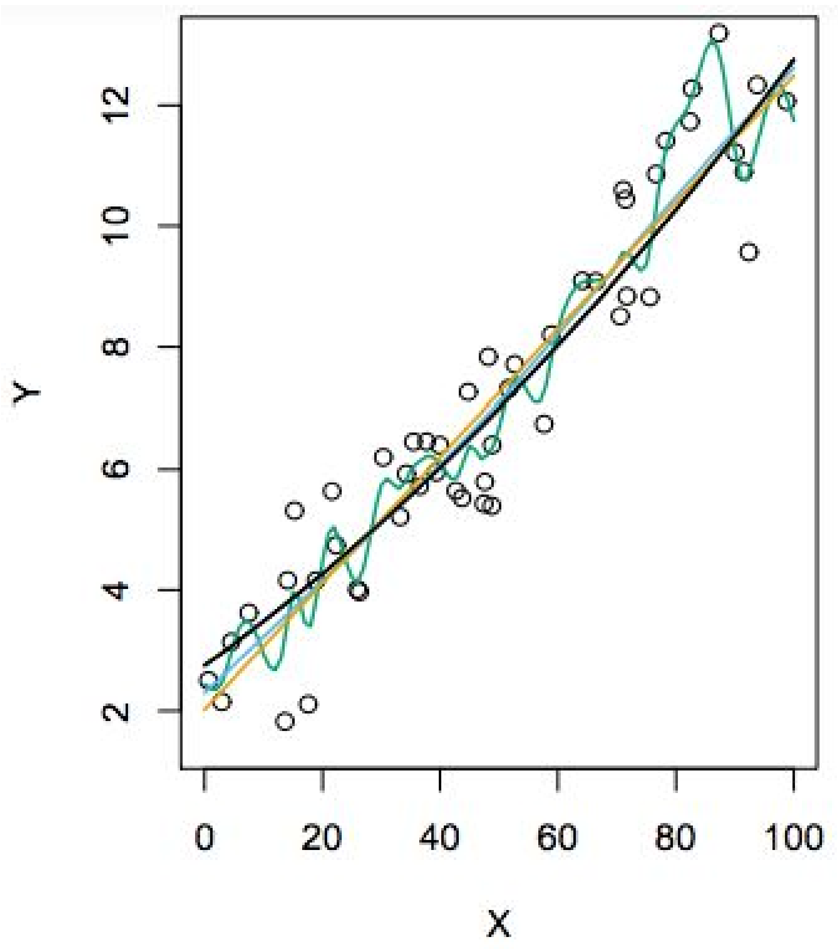
$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{f}(x_i))^2$$

  - (Average squared) difference between prediction and ground truth
- Classification
  - Misclassification rate / classification error rate (what fraction are incorrect?)

$$\frac{1}{n}\sum_{i=1}^{n} I(y_i \neq \hat{y}_i)$$

  - Inverse of accuracy
- Clustering?

Bias-variance trade-off



- Different $\hat{f}(x_i)$ can fit data
- Closer to fit to data?
  - Decrease in bias
  - Increase in flexibility and in variance
- Choosing the flexibility is a trade-off between bias and variance