José Luis Precido Arcila

Theory

- Nonparametric algorithms to estimate stochastic production functions & frontiers.
  - Bayesian
  - Frequentist
- Bayesian estimation of unobserved variables and functions (MCMC).

Industry

- Senior Data Scientist – R&D Innovation Lab @ Austin
  - Revenue opt. – ML algorithms
    – Math. Programming
- Tollway optimization modeler.

Now

- Predictive Defect detection for Steel Industry
- Social Analysis related to workforce rotation. MCMC
- Econometrics
- Wind Energy

Lecture 1: Algorithms and Inference

Statistics:

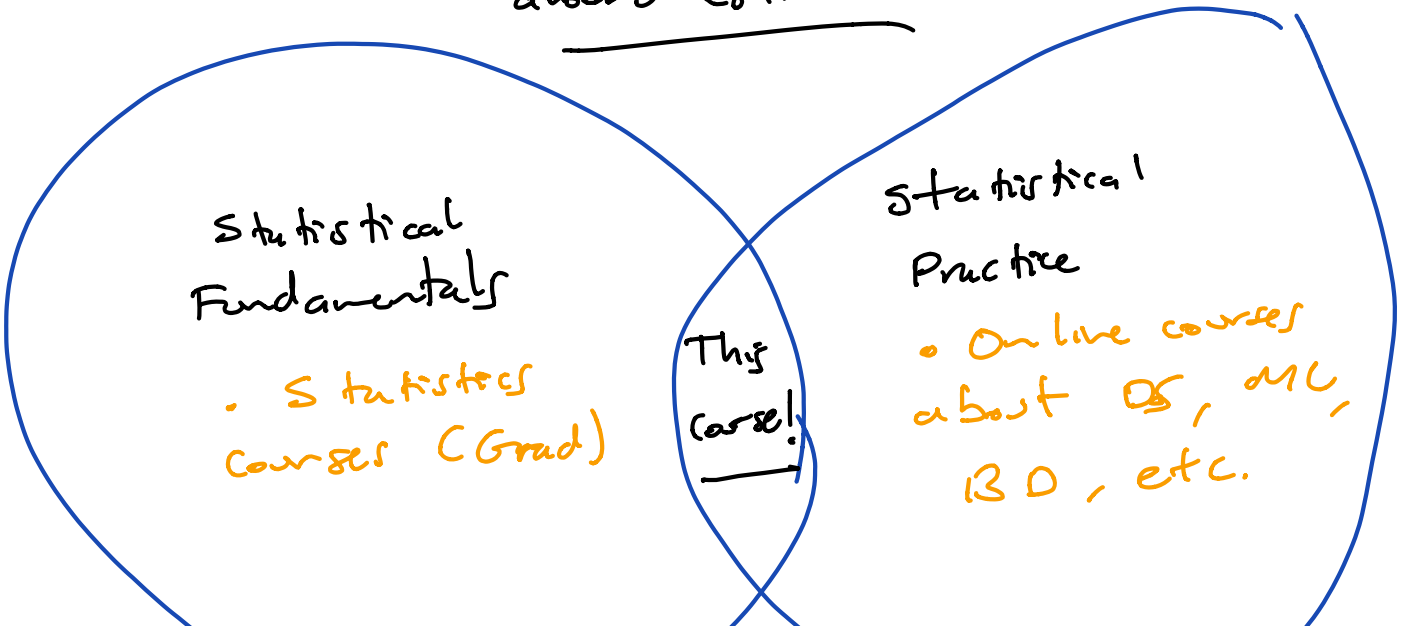
- Branch of math
- Related to prob. theory
- Extracting learnings from data.
- Finding a "true" signal from experience (and separate it from noise).

Algorithmic:

- Like a pipeline that transforms data into more summarized learnings.
- Following a set of data processing steps to produce estimates.

Inferential:

- How "good" is the outcome of the algorithm?
- Measuring the uncertainty around estimates.



Warmup Example

- The mean $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$ summarizes data in to one number.
- How precise is this number?

$\hat{\sigma}_{\bar{x}}$ (standard error) $= \left[\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n(n-1)}\right]^{1/2}$

Denotes estimator

$(\hat{se})$

Sample variance σ^2

$= \left[\frac{\sigma^2}{n}\right]^{1/2} = \frac{\sigma}{\sqrt{n}}$

Inference in regression

- A regression is an estimator of a conditional mean. In same way $\bar{y} = E(y)$

$\hat{y} = E(y|x)$

Height | Gender

For the univariate case

Liver Func. Index $= \beta_0 + \beta_1 \cdot Age$ (model)

$$\rightarrow \quad L_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

L_i: LPI for patient i

x_i: Age of patient i

ε_i: error for observation i.

$$\min_{\beta_0, \beta_1} \sum_{i=1}^{n} (\varepsilon_i)^2 \quad \text{(Algorithm)}$$

Minimize Least Squares Loss Function.

$$\min_{\beta_0, \beta_1} \sum_{i=1}^{n} (L_i - \beta_0 - \beta_1 x_i)^2$$

By the way, your data for this problem, looks like:

L_i	x_i

$$\hat{se}\, LS(1) = \left[\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n(n-2)} \right]^{1/2}$$

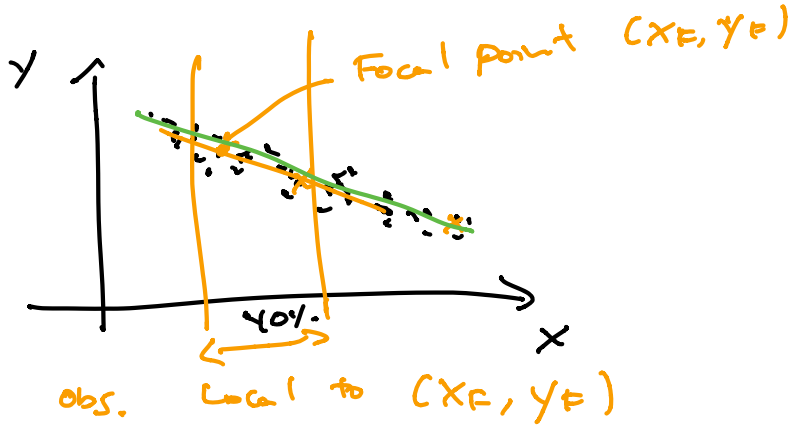
Because we're estimating 2 parameters from data.

Lowess (Local weighted Regression)

- Say we choose $f = 40\%$ of the data to be local to each focal point.

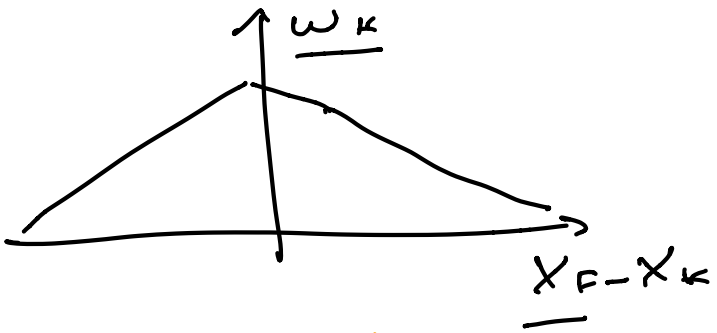
Step 1. Choose the N·f data points closest in X to your focal point (X_F, Y_F)

Your sample



Step 2. Compute a weighted linear regression for the focal point.

This means that behind the weights for each observation there's a weighting pattern.



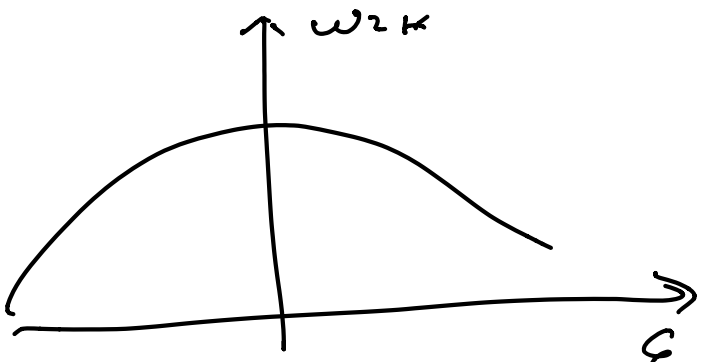
weighting function

where X_K is any other observation.

Step 3. Solve (40%)

$$\min_{\beta_{0i}, \beta_{1i}} \sum_{K=1}^{N \cdot f} (w_K \cdot \varepsilon_K)^2 \quad \text{for } i = 1, \ldots, N$$

Step 4. Apply a second weighting function depending on $\varepsilon_{Fi} = Y_F - \hat{Y}_{F(-i)}$, multiply the weights & redo step 3.



"Give more weight to good predictors from previous step".

$$\varepsilon_i(i) = y_F - \hat{y}_F(i)$$

Step 5. Repeat step 4 until satisfied.

Instead of one regression with N observations we need to compute N · # of iterations regressions with $\lceil N \cdot f \rceil$ observations each.

- There is no exact inference for lowers estimates !!.

How do we get inference then?

→ Bootstrapping → Sampling w/ replacement