Intro to Linear Regression

Note Title 2/29/2016

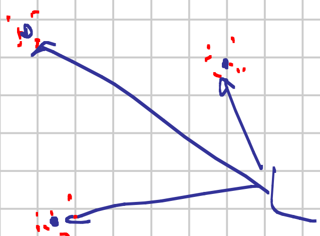
- Midterm
- Assign #4
- Notes
- Intermediate Report
- Move Office Hours Tue @ 2:30-3:30

Input Data: X

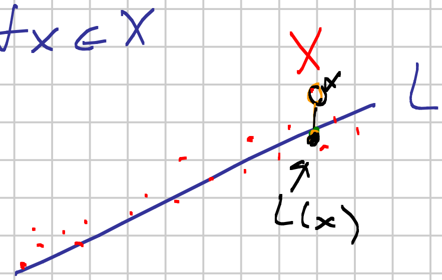
↳ Simple Pattern L

map $\mu : X \to L$

$L(x) \quad \forall x \in X$



centers



Linear Least-Squares Regression

Data → $P \subset \mathbb{R}^2$ $\qquad p = (p_x, p_y) \in P$

$\langle v, u \rangle = \|v\| \cdot \|u\| \cdot \cos(\theta)$

Goal: line $\ell : y = ax + b$ $\qquad P_x = \{ p_x \mid p \in P \}$

parameters $\qquad P_y = \{ p_y \mid p \in P \}$

L_2 cost $\quad L_2(P, a, b) = \sum_{p \in P} (p_y - \underbrace{a p_x - b}_{L(p_x)})^2$ residual

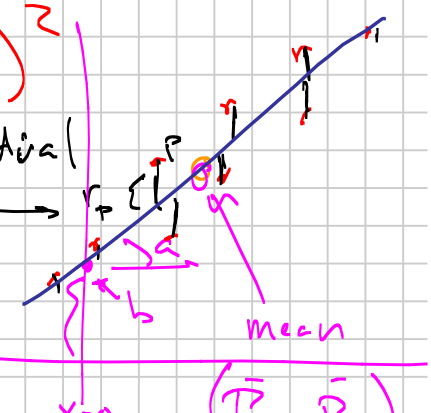
$$\mathrm{Cov}[P_x, P_y] = \frac{1}{n} \sum_{p \in P} (p_x - \bar{P}_x)(p_y - \bar{P}_y)$$

$$\mathrm{Var}[P_x] = \mathrm{Cov}(P_x, P_x)$$

after centering

$$a = \frac{\mathrm{Cov}[P_x, P_y]}{\mathrm{Var}[P_x]} = \frac{\langle P_x, P_y \rangle}{\|P_x\|^2}$$

$$b = \bar{P}_y - a \bar{P}_x$$



$(\bar{P}_x, \bar{P}_y)$

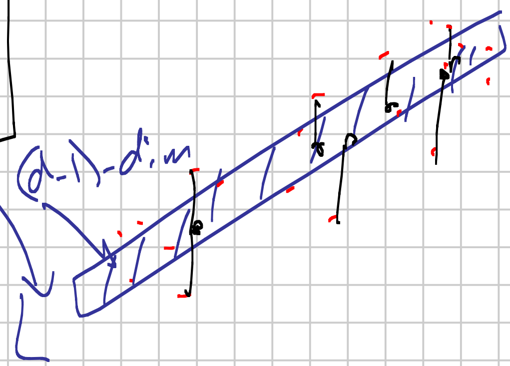
$$\bar{P}_x = \frac{1}{|P|} \sum_{p \in P} p_x$$

$P \subset \mathbb{R}^d \longrightarrow \underset{\substack{n\times d\\ \text{matrix}}}{X}$

$P = (\underbrace{r_1, r_2, r_3 \dots r_{d-1}}_{x}, \underbrace{r_d}_{y})$ $\to P_y$

$$n\left\{ \overbrace{\begin{bmatrix} & & & & & \\ r_1 & r_2 & r_3 & \dots & r_{d-1} & \\ & & & & & \end{bmatrix}}^{d} \right. \begin{bmatrix} a \\ b \end{bmatrix} \approx \begin{bmatrix} P_y \end{bmatrix}$$

$$X$$



$$a = \underbrace{(X^T X)^{-1}}_{\frac{1}{\|P_x\|^2}} \underbrace{X^T P_y}_{\langle P_x, P_y \rangle}$$

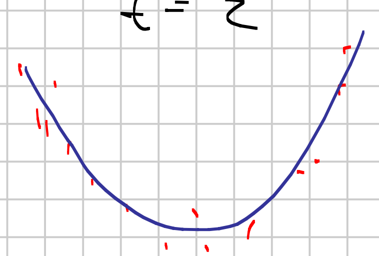
"hat" $\quad H_x = X(X^T X)^{-1} X^T$

$\hat{y} = X_a = H_x y \qquad y = P_y$

Nonlinear (polynomial) LS

fit polynomial degree t $\qquad t = 2$

$$L: \; y = a_0 + a_1 x + a_2 x^2 + \dots a_t x^t$$
$$= \sum_{i=0}^{t} a_i x^i$$



$$P_x \longrightarrow [1, P_x, P_x^2, P_x^3, \dots, P_x^t]$$

"lifting"
"linearization"

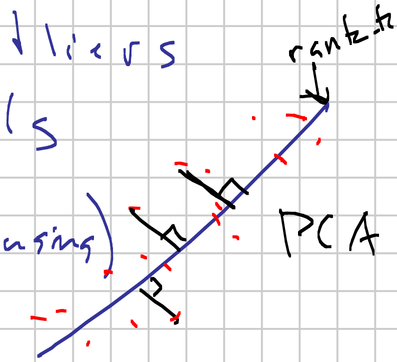
Gauss-Markov Thm

LS sol (above) optimal

$\Rightarrow$ • solution 0 expected error $\quad E[L(x) - x] = 0$

• all errors $r_p = P_y - L(p_x)$ are not known to be correlated

- Squared Error (Min Variance)
  ↳ not resistant to outliers
- Vertical Distance Residuals
- What if $n < d$ (compressed sensing)
- We can add bias



$$\sum_{p \in P}^{!} (p_y - L(p_x))^2 = \text{Var} + \text{Bias}$$

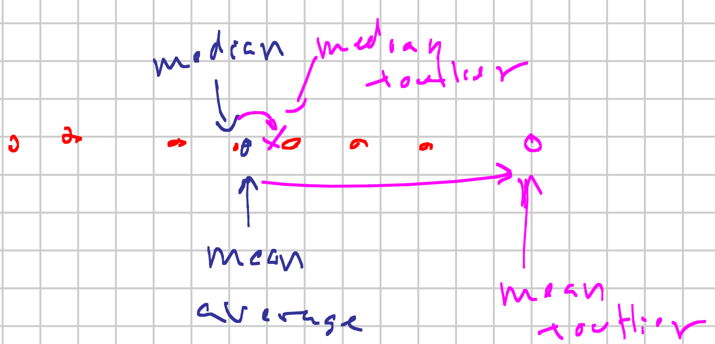
(helps w/ over fitting)

- $(X^T X)^{-1} \longleftarrow$ expensive $\Longrightarrow$ Matrix Sketching

Outliers

Robust (Resistive) Estimator

breakdown point



fraction pts move to ∞ and estimator does not move to ∞

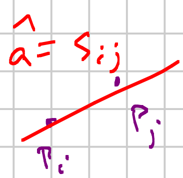
breakdown pt. mean $= \frac{1}{n} \xrightarrow[n \to \infty]{} 0$

" median $= \frac{1}{2}$

Theil-Sen Estimator

bdpt $= 0.293$

(Seigel Est bdpt 0.5)



slopes $s_{ij} = \dfrac{y_j - y_i}{x_j - x_i}$

$P_x \rightarrow$ sort $(P_{x_1} \leq P_{x_2} \leq P_{x_3} \leq \ldots)$

$a = \text{med}(s_{ij} \mid P_{x_i} \leq P_{x_j})$ $\qquad b = \text{med}(\{ y_i - a x_i \})$

Tikhonov Regularization (Ridge Regression)

Assume $\bar{P}_x = 0$ $\bar{P}_y = 0$ centered.

<u>Goal</u> $L_{2,s}(P, a) = \sum_{p \in P} \underbrace{(p_y - a p_x)^2}_{\text{"variance"}} + \underbrace{s a^2}_{\text{regularizer "bias"}}$

Equivalent Formulation

$$L_2^t(P, a) = \sum_{p \in P} (p_y - a p_x)^2 \quad \text{s.t.} \quad a^2 \le t$$

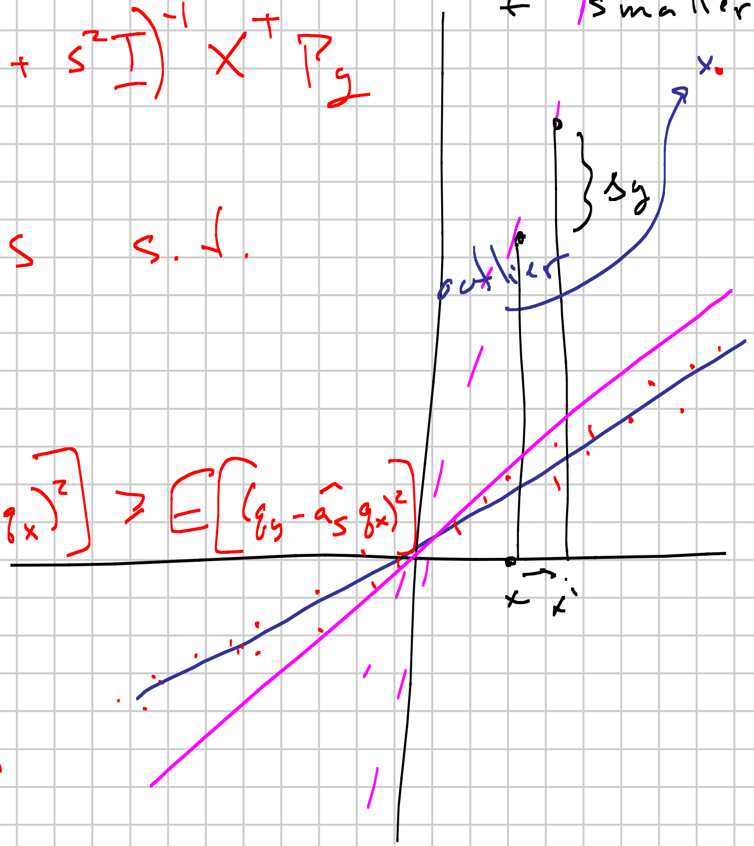
s larger ⟺ t smaller

$$\hat{a}_s = \frac{\langle P_x, P_y \rangle}{\|P_x\|^2 + s^2} = (X^T X + s^2 I)^{-1} X^T P_y$$

$\exists$ some parameter s s.t.

$P \underset{iid}{\sim} D(\theta) \to \hat{a}$, $\hat{a}_s$

$q \sim D(\theta) \to E[(q_y - \hat{a} q_x)^2] \ge E[(q_y - \hat{a}_s q_x)^2]$



How to choose s?

↳ Cross-Validation

Lasso (Basis Pursuit)

$$L_{1,s}(P, a) = \sum_{p \in P} (p_y - a p_x)^2 + s|a|$$

$$L_1^t(P, a) = \sum_{p \in P} (p_y - a p_x)^2 \quad \text{s.t.} \quad |a| \le t$$