Lasso (L_1-Regularized Regression)

Note Title 3/28/2016

Input $P \subset \mathbb{R}^{d+1}$ n points

$P_1, P_2, \dots, P_n$

$P_i = (P_{i1}, P_{i2}, \dots, P_{id}, y_i)$

P_x

$A = [a_0, a_1, a_2, \dots, a_d]$ $\quad a_0 = \begin{bmatrix} \\ \\ \\ \end{bmatrix}$

only important unit in $(d+1)$-coord $= y$

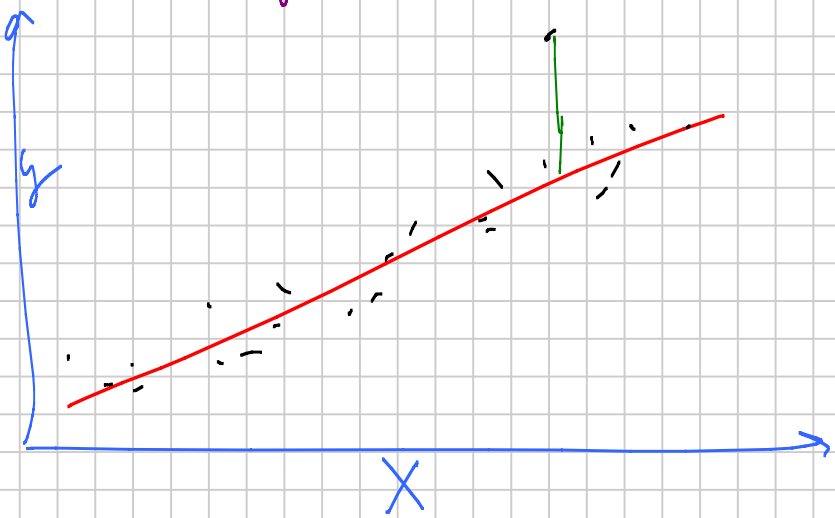
$$\sum_{P_i \in P} \left(f_A(P_i) - y_i\right)^2$$

$$X = [1 \; P_x]$$

$$\min_A \|XA - y\|_2 \implies A = (X^T X)^{-1} X^T y$$

Regularization

- Bayesian Prior
- "Regression to the mean"



Midterm Test (T-F) 75% 86%

- Know 50 Qs
- Guess 50 Qs

E[score] = 75%

Take top 10 student

Ave[score] = 80%

Final

Expected of top students?

↳ E[top] = 75%

Tikhonov Regularization (Ridge Regression)

$$\hat{A} = \arg\min_A \left(\|XA - y\|_2 + s\|A\|_2 \right)$$

$$\hat{A} = (X^T X + s\mathbb{I})^{-1} X^T y$$

$$\|A\|_2 = \sum_{i=0}^{d} a_i^2$$

$$\|A\|_1 = \sum_{i=0}^{d} |a_i|$$

Lasso

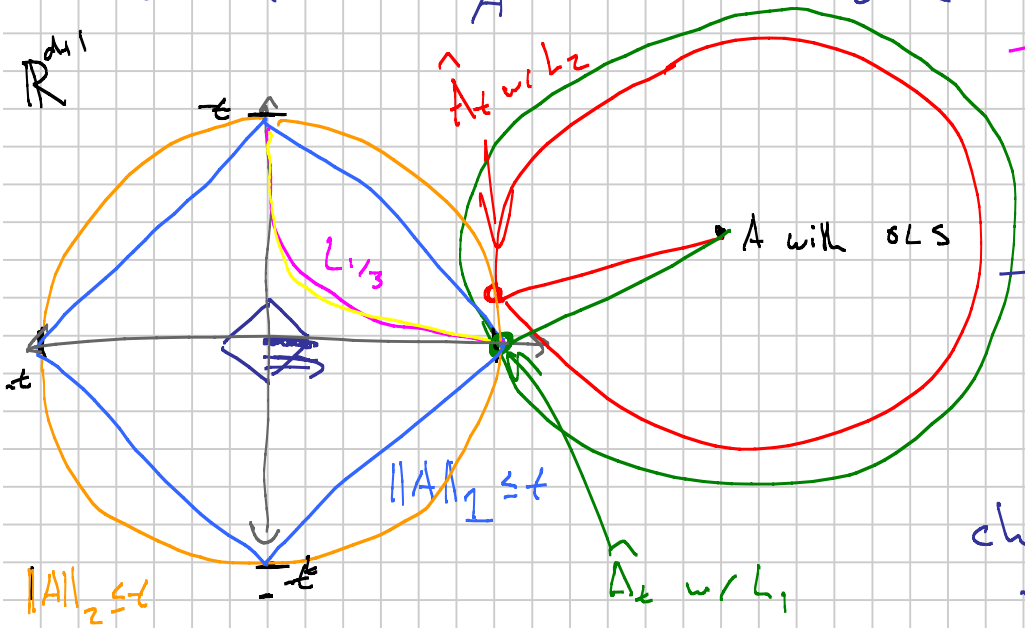
$$\hat{A}_s = \arg\min_A \left(\|XA - y\|_2 + s\|A\|_1 \right)$$

- in high dim $\Rightarrow$ bias towards A sparse
- force to consider value s.

$\forall s \Rightarrow \exists t$ s.t. $\quad t_s = \|\hat{A}_s\|_1$

s increase
t decrease

$$\hat{A}_s = \hat{A}_t = \arg\min_A \|XA - y\|_2 \quad \text{s.t.} \quad \|A\|_1 \le t$$



t large
↳ OLS A dense

t very small
↳ very sparse
↳ 1 non-zero

choose coord j_1

$$j_1 = \arg\max_j |\langle X_j, y \rangle|$$

$$a_{j_1}(t) = t$$

$$r(t) = y - \sum_{j=0}^{d} X_j a_j(t)$$

$$a_{j_1}(t) = a_{j_1}(t_2) + (t - t_2) b_1$$
$$a_{j_2}(t) = (t - t_2) b_2$$

until

$$|\langle X_{j_1}, r(t) \rangle| = |\langle X_{j_2}, r(t) \rangle|$$

- Values of a_j can be negative
- as t increases, value of a_j can oscilate between neg and pos.
- If a_j hits 0, snap.
  ↳ only change a_j if $|\langle \cdot, \cdot \rangle| = K, \cdot|$

Is this better than OLS?

generalization error