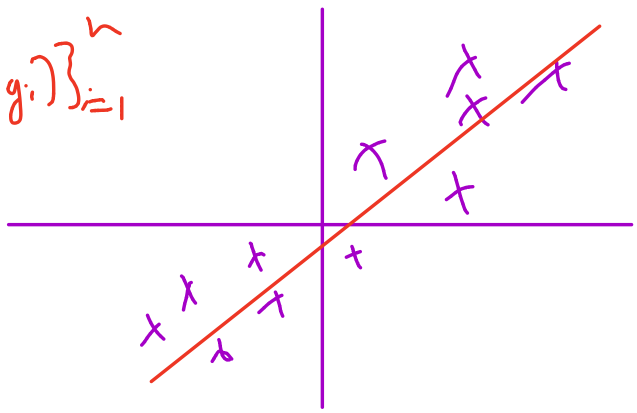
Regression:

$T = \{(x_i, y_i)\}_{i=1}^{n}$



Under the OLS model

$$Y = f(x) + \varepsilon = X^T\beta + \varepsilon$$

$$\hat{\beta}_{OLS} = \underset{\beta}{\operatorname{argmin}} \; (Y - X\beta)^T (Y - X\beta)$$

$$\hat{\beta}_{Ridge} = \underset{\beta}{\operatorname{argmin}} \; (Y - X\beta)^T (Y - X\beta) + \lambda \|\beta\|_2^2$$

$$\boxed{\hat{\beta}_{OLS} = (X^T X)^{-1} X^T Y}$$

$$\boxed{\hat{\beta}_{Ridge} = (X^T X + \lambda I)^{-1} X^T Y}$$

Lasso Regression

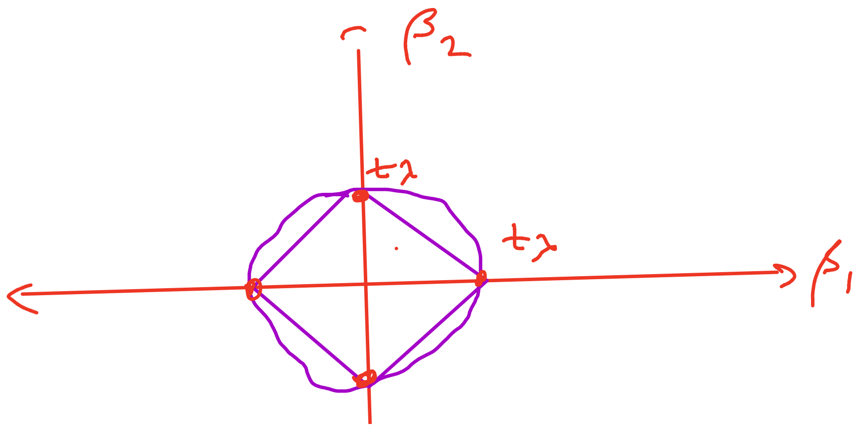
$$\hat{\beta}_{Lasso} = \underset{\beta}{\operatorname{argmin}} \; (Y - X\beta)^T (Y - X\beta) + \lambda \|\beta\|_1$$

equivalent to the minimization problem

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \; (Y - X\beta)^T (Y - X\beta)$$

subject to $\sum_{j=1}^{p} |\beta_j| \leq t_\lambda$

Solving this Legrande system



So Lasso is a more restrictive class of models.

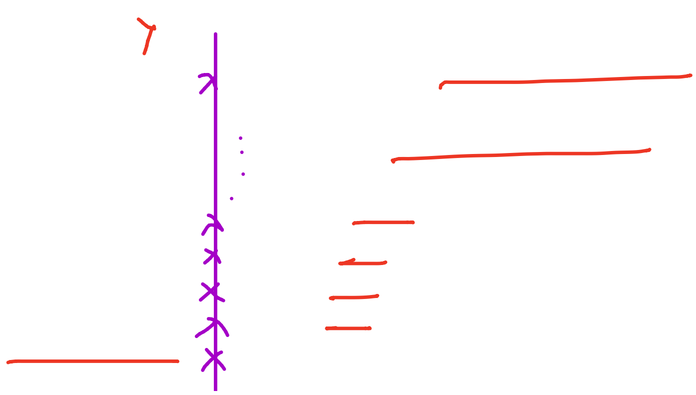
<u>Ex:</u> Columns of X are orthonormal

$$\hat{\beta}_{Lasso} = \text{sgn}(\hat{\beta}_j)\left(|\hat{\beta}_j| - \lambda\right)_+$$

$$(a)_+ = \begin{cases} a & a \geq 0 \\ 0 & a < 0 \end{cases}$$

Classification

Goal: Estimate (X, Y)



Bayes Problem:

Assume we know $p(x,y)$ and we look to formulate the best estimate y from x

$$\hat{f} = \underset{f}{\arg\min}\; EPE(f)$$

arb. loss

$$= \underset{f}{\arg\min}\; \mathbb{E}_T\left[L(Y, f(X))\right]$$

$$= \underset{f}{\arg\min}\; \mathbb{E}_X\left\{\underbrace{\mathbb{E}_{Y|X}\left[L(Y, f(x)) \mid X\right]}\right\}$$

Idea: minimize for each value of x. So

$$\hat{f}(x) = \underset{f}{\operatorname{argmin}}\ \mathbb{E}_{Y|X}\left(y - f(x)\right)^2$$

$$= \underset{c}{\operatorname{argmin}}\ \mathbb{E}_{Y|X}\left(Y - c\right)^2$$

$$= \mathbb{E}_{Y|X}(y)$$