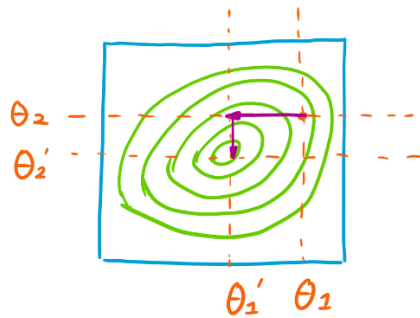


Optimization for Linear Models with Regularization

Zengchang Qin (PhD)

1) Coordinate Descent



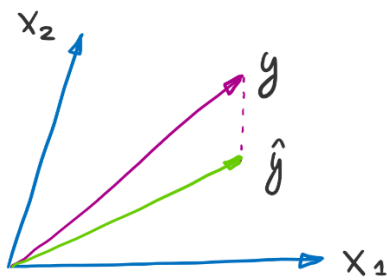
$$J(\theta) = \frac{1}{2} \sum_{i=1}^n (x_i^T \theta - y_i)^2 + \lambda \|\theta\|_1$$

$$\|\theta\|_1 = \sum_j |\theta_j|$$

$$\theta_j : j = 1, 2, \dots, n$$

$$(\theta_1, \theta_2) \rightarrow (\theta_1', \theta_2')$$

2) Forward Selection

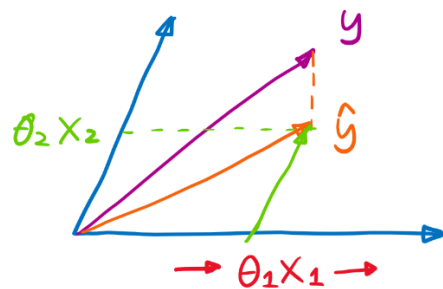


$$\hat{y} = \theta_0 + \theta_1 x_1 + \theta_2 x_2$$

$$y - \hat{y} \perp L(x_1, x_2)$$

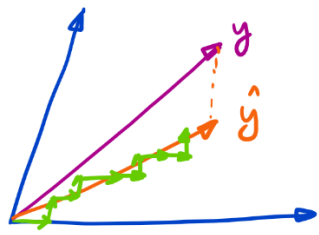
$$X = \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix} \quad \theta = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \quad Y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

$$\begin{pmatrix} x_{11} \\ x_{21} \end{pmatrix} \theta_1 + \begin{pmatrix} x_{12} \\ x_{22} \end{pmatrix} \theta_2 \approx \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

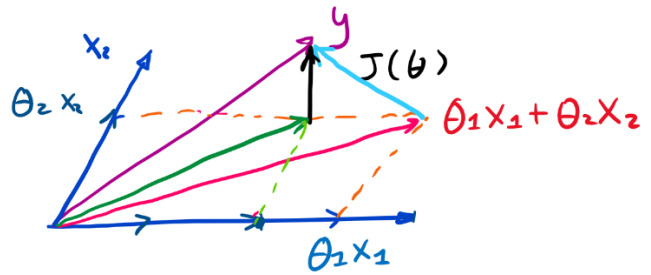


$$J(\theta) = \|y - \hat{y}\|_2^2$$

4) Forward Strategies

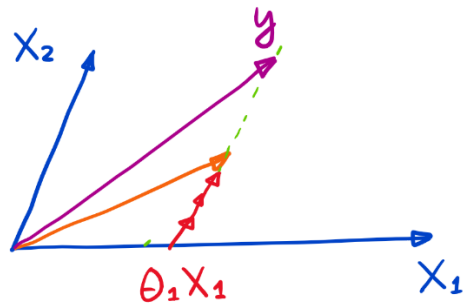
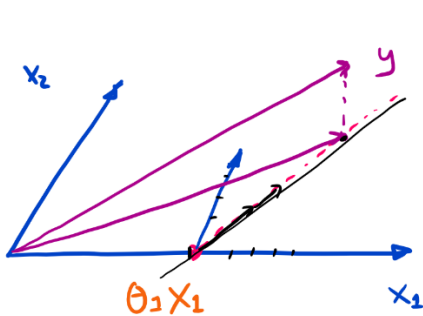


$$\begin{cases} \theta_1 = k\varepsilon \\ \theta_2 = m\varepsilon \end{cases}$$



$$\theta_1 x_1 + \theta_2 x_2 = \hat{y}$$

4) Least Angle Regression (LARS)



$$\begin{aligned} \theta_1 x_1 + \theta_2 (x_1 + x_2) &= (\theta_1 + \theta_2) x_1 + \theta_2 x_2 \\ &= \theta'_1 x_1 + \theta'_2 x_2 \end{aligned}$$