Regularization

Overfit models tend to have large parameters.

Minimize Loss function

$$\begin{cases} \min_{\beta} L(\beta) \\ \text{s.t.} \quad \sum |\beta_i| \leq \eta \end{cases}$$

$$\min_{\beta} L(\underline{\beta}) + \lambda \sum |\beta_i|$$

Lin Reg. $L(\beta) = (y_{true} - y_{pred})^2$

Log. Reg $L(\beta) = -MLE$

$$\underline{x}^{(k+1)} = \underline{x}^{(k)} - \alpha g(\underline{x})$$

hyperparameter

$$\min_{\beta} L(\beta) + \lambda \|\beta\|_k$$

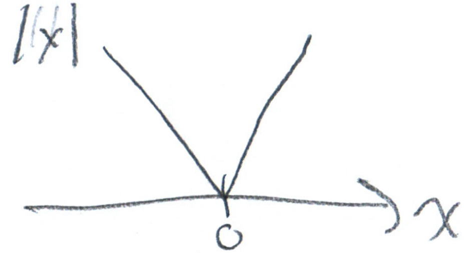
$k=1$ LASSO

$k=0$, 0-norm

$\|\beta\|_0 = \#$ of $\beta_i \neq 0$

too hard (NP)

$k=1$, 1-norm $|\beta_i|$



Sparse solutions

lots of $\beta_i = 0$

$k=2$, 2-norm β_i^2

Unique solution

Cont. deriv.

lots of small β_i

(few $\beta_i = 0$)

hard penalty on large β_i

$$\min_{\beta} L(\beta) + \lambda_1 \|\beta\|_1 + \lambda_2 \|\beta\|_2$$

Elastic Net Regularization

15: Geometry

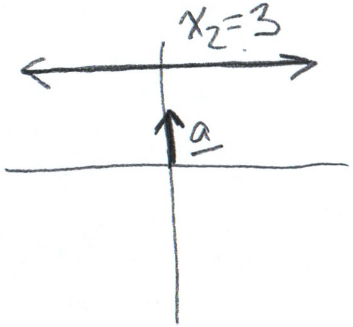
Slope intercept: $y = mx + b$

$$x_2 = m x_1 + b$$

Standard form: $a_1 x_1 + a_2 x_2 = b$

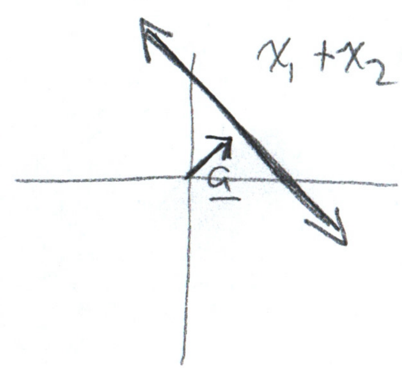
$$\underline{a} \cdot \underline{x} = b, \quad \underline{a} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$$

$$\underline{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$



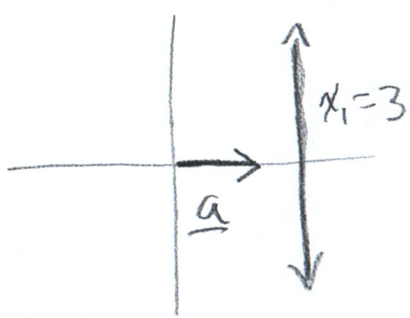
$$0x_1 + 1x_2 = 3$$

$$\underbrace{\begin{pmatrix} 0 \\ 1 \end{pmatrix}}_{\underline{a}} \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 3$$



$$1x_1 + 1x_2 = 5$$

$$\underbrace{\begin{pmatrix} 1 \\ 1 \end{pmatrix}}_{\underline{a}} \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 5$$



$$1x_1 + 0x_2 = 3$$

$$\underbrace{\begin{pmatrix} 1 \\ 0 \end{pmatrix}}_{\underline{a}} \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 3$$

$\underline{a}$ is normal to the line/plane/hyperplane