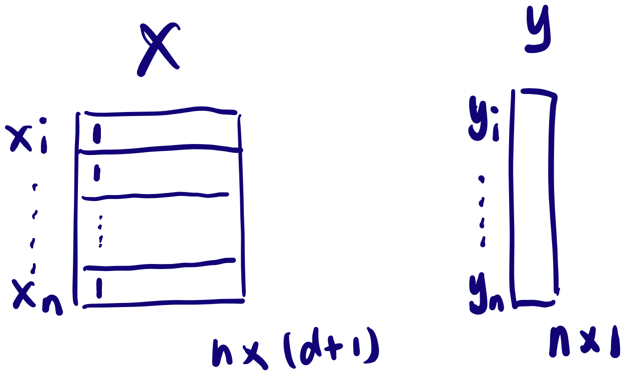
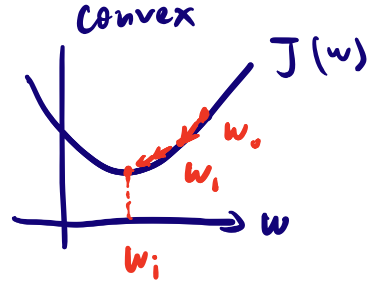
Sep 19, 2022. Mon

$$\hat{W}_{MLE} = (X^T X)^{-1} X^T y$$



$$J(w) = \frac{1}{2} \sum_{i=1}^{N} (y_i - w^T x_i)^2$$

convex



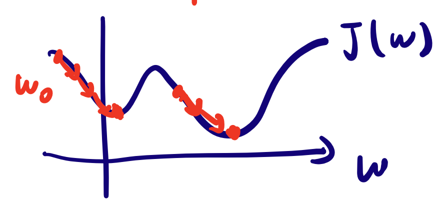
Gradient Descent

initialize w_0

$$\frac{\partial J(w)}{\partial w} = \sum_{i=1}^{N} (y_i - w^T x_i) \cdot x_i$$

$$w_{i+1} = w_i - \eta \frac{J(w_i)}{\partial w_i}$$

until covergence



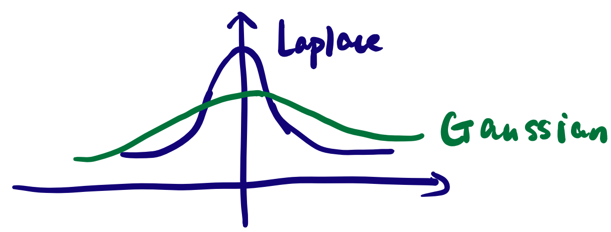
non-convex

Robust Regression.

$y \sim \text{Laplace}(w^T x, b)$

$$\begin{cases} y = w^T x + \varepsilon \\ \varepsilon \sim \text{Laplace}(0, b) \end{cases}$$

$$p(m) = \frac{1}{2b} \exp\left(- \frac{|m - \mu|}{b}\right)$$



probabilistic interpretation

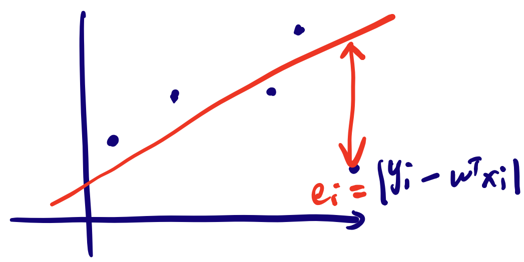
$$LL(w) = \log \prod_{i=1}^{N} p(y_i \mid w, b)$$

$$= \log \prod_{i=1}^{N} \frac{1}{2b} \exp\left(-\frac{|y_i - w^T x_i|}{b}\right)$$

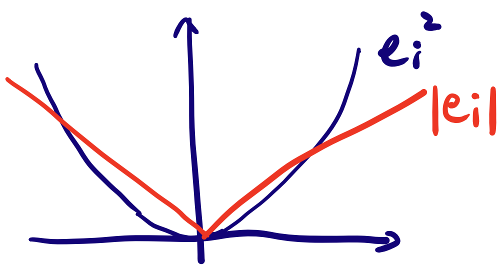
$$= \log \frac{N}{2b} - \frac{1}{b} \sum_{i=1}^{N} \underbrace{|y_i - w^T x_i|}_{\text{error}}$$

Geometric interpretation

$$J(w) = \sum_{i=1}^{N} |y_i - w^T x_i|$$



$$\text{OLE} \quad J(w) = \sum_{i=1}^{N} (y_i - w^T x_i)^2$$



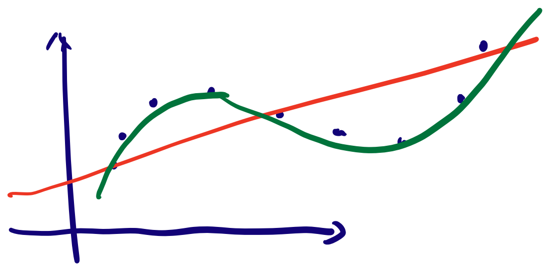
$e_i = y_i - w^T x_i$

$$y = w^T x$$

basis function $\phi(x) = [1, x, x^2, \cdots x^d]$

$$y = w_0 + w_1 x + w_2 x^2 + \cdots + w_d x^d$$

linear to w, non-linear to x



$\phi(x) = [1, x_1, x_2, x_3, x_1x_2, x_1x_3, x_2x_3, \cdots]$

Ridge Regression

$$\theta(w) = J(w) + \underbrace{\lambda \|w\|_2^2}_{l_2 \text{ regularization}}$$

l_2 norm $\|w\|_2^2 = w_1^2 + w_2^2 + \cdots + w_d^2$

$\|w\|_2 = (w_1^2 + w_2^2 + \cdots + w_d^2)^{\frac{1}{2}}$

$$\theta(w) = \sum_{i=1}^{N} (y_i - w^T x_i)^2 + \lambda \|w\|_2^2$$

$$= (y - Xw)^T (y - Xw) + \boxed{\lambda w^T w}$$

hyperparameter

$$\frac{\partial \theta(w)}{\partial w} = 0 \qquad w = (X^T X + \lambda I)^{-1} X^T y$$

Correlated variables

$X = [x_1, x_2] \qquad x_2 = x_1 + \epsilon$

$y = w_0 + w_1 x_1 + w_2 x_2$ unstable

$y = w_0 + 2w_1 x_1 + 0 x_2$

$y = w_0 + 0 x_1 + 2w_2 x_2$

$\vdots$

$$\|w\|_2^2 \Rightarrow y = w_0 + w_1 x_1 + w_2 x_2$$

l_p norm

$$\|w\|_p = (w_1^p + w_2^p + \cdots w_d^p)^{\frac{1}{p}}$$

l_2 norm $\|w\|_2 = (w_1^2 + w_2^2 + \cdots w_2^2)^{\frac{1}{2}}$

l_∞ norm $\|w\|_\infty = \max(|w_1|, \cdots, |w_d|)$

l_1 norm $\|w\|_1 = |w_1| + |w_2| + \cdots |w_d|$

l_0 norm number of nonzero in w

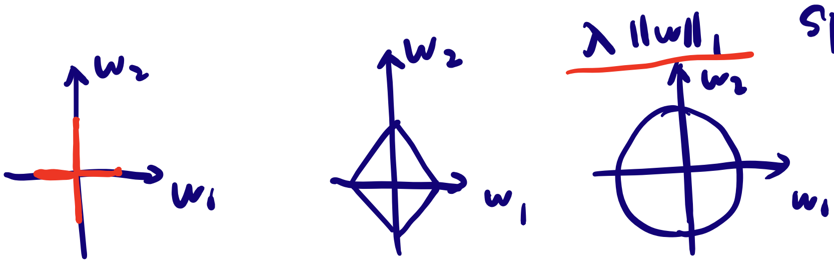
$$J(w) = \sum_{i=1}^{N} (y_i - w^T x_i)^2 + \lambda \|w\|_0 \qquad \text{LASSO}$$

$$\Downarrow$$

$$\lambda \|w\|_1$$

Sparsity in w

prior on w following Laplace



$p=0$ l_0 norm $\qquad p=1$ $\qquad (w_1^2 + w_2^2)^{\frac{1}{2}}$

$|w_1| + |w_2|$