Bayesian Regression

$$y \mid x \sim \mathcal{N}(w^T x, \sigma^2)$$

$$p(y_i \mid x_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y_i - w^T x_i)^2}{2\sigma^2}\right]$$

$$L(D) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y_i - w^T x_i)^2}{2\sigma^2}\right]$$

$$LL(D) = \left[\sum_{i=1}^{N} \log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \sum_{i=1}^{N} \frac{1}{2\sigma^2}(y_i - w^T x_i)^2\right]$$

We want to find σ^2, w that maximize $LL(D)$

OLE

$$y = w^T x$$

x_1	y_1
x_2	y_2
$\vdots$	$\vdots$
x_N	y_N

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

$$w = (X^T X)^{-1} X^T y$$

---

w: find w that maximizes $\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i - w^T x_i)^2\right]$

Same as $\underset{w}{\text{minimize}} \ \frac{1}{2}\sum (y_i - w^T x_i)^2$

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

$$\hat{\sigma}^2_{MLE} = \cancel{\frac{1}{N}(y - Xw)} \ \frac{1}{N}(y - Xw)^T (y - Xw)$$

$\hat{w}_{MAP}$ $\qquad$ $w_{Bayesian}$

Robust Regression

$$J(w) = \frac{1}{2} \sum (y_i - w^T x_i)^2 \Rightarrow \frac{1}{2} \sum |y_i - w^T x_i|$$

$$LL(D) = \_\_\_ - \sum_{i=1}^{N} \underline{\frac{1}{2\sigma^2} (y_i - w^T x_i)^2}$$

Replace with another distribution

$|y_i - w^T x_i|$

Laplace

Student-t

Why is Linear Regression Susceptible to outliers:

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

$$LL(w) = \sum \log(\text{pdf})$$

$$\frac{\cancel{(w^T x_i - \mu)}^2 \; (y_i - w^T x_i)^2}{2\sigma^2}$$

Fixing this:

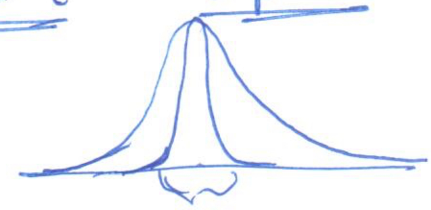
$$J(w) = \frac{1}{2}\sum_{i=1}^{N} |y_i - w^T x_i|$$

least abs. deviation

Use a different dist.

$$y_i \sim \mathcal{N}(y_i \mid w^T x_i, \sigma^2)$$

Student-t   Laplace.



Prior $p(w) \rightarrow \mathcal{N}(w \mid 0, \tau^2 I)$

$$\begin{pmatrix} \tau^2 & & 0 \\ 0 & \tau^2 & \\ & & \tau^2 \\ & & & \ddots \end{pmatrix}$$

$$p(w \mid D) = \frac{p(w)\prod_{i=1}^{N} p(y_i \mid w)}{\int p(w)\prod_{i=1}^{N} p(y_i \mid w)}$$

Posterior

$$p(x_i \mid w) \sim \mathcal{N}(w^T x_i, \sigma^2)$$

$$= \mathcal{N}\Big(w \,\Big|\, \underbrace{\left(X^T X + \frac{\sigma^2}{\tau^2} I\right)^{-1} X^T y}_{\text{mean}},\ \underbrace{\cancel{\sigma^2}\left(X^T X + \frac{\sigma^2}{\tau^2} I_N\right)^{-1}}_{\text{Covariance}}\Big)$$

MLE: $w_{MLE} = \left((X^T X^{-1}) X^T y, (X^T X)^{-1} \right)$

Ridge $w_{Ridge} = (X^T X + \lambda I)^{-1} X^T y$

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

Generalized Linear Models. (GLM)

y is connected to $w^T x$ $y \sim \mathcal{N}(y \mid w^T x, \sigma^2)$

$y \sim \text{Bernoulli}(y \mid \sigma(w^T x))$ → (Logistic Regression)

$\sigma(w^T x)$ → Sigmoid