

# 1. LINEAR REGRESSION

$$MSE = E[(f(x) - y)^2]$$

let  $f(x) = \theta$

$$MSE = E[(\theta - y)^2]$$

Using Law of Iterative Expectations.

$$E[(\theta - y)^2] = E[E[(\theta - y)^2 | X]]$$

Find  $\theta^*$  that minimizes  $MSE$  <sup>conditional</sup> for all  $X$  ( $E[(\theta - y)^2 | X]$ )  $\Rightarrow$  It will also minimize  $MSE$

Take derivative with respect to  $\theta$  and set it = 0.

$$\Rightarrow \frac{\partial E[(\theta - y)^2 | X]}{\partial \theta} = E\left[\frac{\partial(\theta - y)^2}{\partial \theta} | X\right]$$

$$= E[2(\theta - y)(1) | X]$$

$$= 2E[(\theta - y) | X] = 0$$

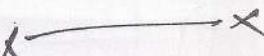
$$\Rightarrow E[(\theta - y) | X] = 0$$

$$\boxed{f(x) = \theta} \text{ substituting back.}$$

$$\Rightarrow E[f(x) | X] - E[y | X] = 0$$

$$\Rightarrow E[f(x) | X] = E[y | X]$$

$f(x)$  is a function of  $X$  and expectation is for fixed values of  $X$   
therefore  $E[f(x) | X] = f^*(x)$

$$\boxed{f^*(x) = E(y | X)}$$


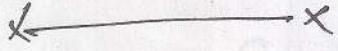
# Gaussian Noise Regression Model

$$a) p(t|X, \omega, \beta) = \prod_{n=1}^N N(t_n | \omega^\top \phi(x_n), \beta_n^{-1}) \\ = \prod_{n=1}^N \frac{1}{\sqrt{2\pi\beta_n^{-1}}} \exp \left\{ -\frac{(t_n - \omega^\top \phi(x_n))^2}{2\beta_n^{-1}} \right\}$$

The log-likelihood.

$$\log(p(t|X, \omega, \beta)) = \log \left[ \prod_{n=1}^N \frac{\sqrt{\beta_n}}{\sqrt{2\pi}} \exp \left\{ -\frac{\beta_n}{2} (t_n - \omega^\top \phi(x_n))^2 \right\} \right] \\ = \sum_{n=1}^N \left[ \frac{1}{2} \log \beta_n - \frac{1}{2} \log(2\pi) - \frac{\beta_n}{2} (t_n - \omega^\top \phi(x_n))^2 \right] \\ = \frac{N}{2} \log \beta_n - \frac{N}{2} \log(2\pi) - \frac{\beta_n}{2} \sum_{n=1}^N (t_n - \omega^\top \phi(x_n))^2 \\ \log(p(t|X, \omega, \beta)) = \frac{N}{2} \log \beta_n - \frac{N}{2} \log(2\pi) - \frac{\beta_n}{2} \sum_{n=1}^N (t_n - \omega^\top \phi(x_n))^2$$

b). Sum of squared error is maximum likelihood under a Gaussian Noise Model. Here it is  $\beta_n$  times Sum of Squared Error.



# Weighted Linear Regression

$$a) P(y|X, \beta, \sigma^2) = \prod_{i=1}^N \mathcal{N}(y_i | u_i^\top \beta, \sigma_i^2)$$

$$\begin{aligned} \log(P(y|X, \beta, \sigma^2)) &= \log \left[ \prod_{i=1}^N \left[ \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp \left[ \frac{(-y_i - u_i^\top \beta)^2}{2\sigma_i^2} \right] \right] \right] \\ &= \sum_{i=1}^N \left[ -\log \sqrt{2\pi\sigma_i^2} - \frac{1}{2\sigma_i^2} (y_i - u_i^\top \beta)^2 \right] \\ &= \sum_{i=1}^N \left[ -\frac{1}{2} \log(2\pi) + \frac{1}{2} \log(\sigma_i^2) - \frac{1}{2\sigma_i^2} (y_i - u_i^\top \beta)^2 \right] \\ &= -\frac{N}{2} \log(2\pi) + \frac{N}{2} \log \sigma_i^2 - \sum_{i=1}^N \frac{1}{2\sigma_i^2} (y_i - u_i^\top \beta)^2 \end{aligned}$$

$$\boxed{\hat{\beta}^* = \underset{\beta}{\operatorname{argmax}} \left[ \sum_{i=1}^N \frac{1}{2\sigma_i^2} (y_i - u_i^\top \beta)^2 \right]}$$

b) Let  $w_i = \frac{1}{\sigma_i^2}$  as they are inversely proportioned.

Small error  $\rightarrow$  large weight  
Large error  $\rightarrow$  small weight.

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

$$\nabla \log P(y|X, \beta, \sigma^2) = \sum_{i=1}^N \frac{\partial w_i}{\partial \beta} (y_i - x_i^T \beta) / (\sigma_i^2)$$

$$0^T = \sum_{i=1}^N c_i x_i^T y_i - \beta^T \sum_{i=1}^N w_i x_i x_i^T$$

Taking derivative and setting it = 0.

$$(0^T)^T = (Y^T W^T X - \beta^T W^T X^T X)^T$$

Writing in vector form and taking Transpose on both sides.

$$0 = Y W X^T - \beta W X X^T$$

$$\beta W X X^T = Y W X^T$$

$$\boxed{\beta = (Y W X^T)(W X X^T)^{-1}}$$

$$X \quad \quad \quad X$$

## 2. Regularization.

a) We cannot compute according to (3) because  $M \gg N$  which makes  $(X^T X)^{-1}$  not invertible.

b)  $J_R(\beta) = (X\beta - y)^T (X\beta - y) + \lambda \|\beta\|^2$   
 $= (X\beta - y)^T (X\beta - y) + \lambda I \beta^T \beta$

Take partial derivative with respect to  $\beta$  and set it = 0  
 $\nabla J_R(\beta) = 0$

$$0 = 2X^T(X\beta - y) + 2\lambda \beta$$

$$0 = X^T X \beta - X^T y + \lambda I \beta$$

$$X^T y = X^T X \beta + \lambda I \beta$$

$$X^T y = \beta (X^T X + \lambda I)$$

$$\boxed{\beta^* = (X^T X + \lambda I)^{-1} X^T y}$$

$$b = \frac{b}{1-b} (1-b) \quad \leftarrow \rightarrow$$

$$\frac{1-b}{b}$$

$$\text{pol}\left(\frac{1-b}{b}\right) = -10 + 5 * x$$

$$\text{pol}(0.77) = -10 + 5 * x$$

### 3. Classification

Logistic regression

$$\log(\text{odds}) = -10 + 2 \cdot \mu_1$$

a)

$$\log\left(\frac{P}{1-P}\right) = -10 + 2 \cdot \mu_1$$

$$e^{\log(P/(1-P))} = e^{-10+2\mu_1}$$

$$\frac{P}{1-P} = e^{-10+2\mu_1}$$

$$P = e^2(1-P)$$

$$P = e^2 - e^2 P$$

$$P + e^2 P = e^2$$

$$P(1+e^2) = e^2$$

$$P = \frac{e^2}{1+e^2}$$

$P = 0.88$  is the probability of a cell being unaffected with diameter = 6.

b)

$$\log\left(\frac{0.9}{1-0.9}\right) = -10 + 2\mu_1$$

$$\frac{\log\left(\frac{0.9}{1-0.9}\right) + 10}{2} = \mu_1$$

$$\mu_1 = 6.098$$

## Softmax for Multi-Class Classification

a). Activation Functions.

$$\text{Type A} = 2u_1 + 5u_2 + 5$$

$$\text{Type B} = 5u_1 + 10u_2 + 1.5$$

$$\text{Type C} = 5u_1 + 2u_2 + 1$$

Class probabilities.

$$P(C_A|u) = \frac{\exp(2u_1 + 5u_2 + 5)}{\exp[(2u_1 + 5u_2 + 5) + (5u_1 + 10u_2 + 1.5) + (5u_1 + 2u_2 + 1)]}$$

$$P(C_B|u) = \frac{\exp(5u_1 + 10u_2 + 1.5)}{\exp[(2u_1 + 5u_2 + 5) + (5u_1 + 10u_2 + 1.5) + (5u_1 + 2u_2 + 1)]}$$

$$P(C_C|u) = \frac{\exp(5u_1 + 2u_2 + 1)}{\exp[(2u_1 + 5u_2 + 5) + (5u_1 + 10u_2 + 1.5) + (5u_1 + 2u_2 + 1)]}$$

b)  $u_1 = 10, u_2 = 2 \Rightarrow$  Plugging values in above equation:

$$P(C_A|u_1, u_2) = 1.4 \times 10^{-16}$$

$$P(C_B|u_1, u_2) = 0.999$$

$$P(C_C|u_1, u_2) = 6.825 \times 10^{-8}$$

Predicted type is Type B as the class probability of Type B is highest that is  $\boxed{0.999}$ .