X

1	2	3
4	5	6

$\cancel{X}\theta$

9

10

11

$X\theta$

$\begin{pmatrix} h(x_1) \\ h(x_2) \end{pmatrix}$

$$z^T z = \sum_i z_i^2$$

$$X\theta - \bar{y} = \uparrow \quad \downarrow$$

$$(X^\theta - \bar{y}) \quad \begin{pmatrix} 5 \\ 6 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \cancel{(5\ 6)}$$

$$\begin{pmatrix} 5 & 6 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = 5 \quad 6$$

$$\cancel{\frac{1}{2}} \sum_{i=1}^{m} \frac{1}{2} w^i \left(\theta^T x^{(i)} - y^{(i)}\right)^2 = \sum_{i=1}^{m} (h_\theta(x^i) - y^i) \times \frac{1}{2} w_i \times (h_\theta(x^i) - y^i)$$

$\uparrow$

$h_\theta(x^{(i)}) - y^i$ $=$

$\uparrow$

$A = \theta^T \quad B = X^T$

$C = \theta$

$$(\theta^T X - \bar{y})^T (\theta^T X - \bar{y}) = \sum_{i=1}^{m} (\theta^T x^i - y^i)^2$$

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} =$$

$X\theta\theta$

$\Rightarrow$

and $\sum a_i x_i x_i =$

$= \begin{pmatrix} a_i & \\ & a_i \end{pmatrix} X^T X$

$$\underbrace{(X\theta - \bar{y})^T}_{} \times W \times (X\theta - \bar{y})$$

$((h(x^1) - y^1) \times w^1, (h(x^2) - y^2) w^2, \ldots$

$$W = \begin{pmatrix} \frac{w^1}{2} & & \\ & \frac{w^2}{2} & \\ & & \ddots \; \frac{w^m}{2} \end{pmatrix}$$

$$\nabla_\theta \operatorname{tr} \theta X^T X \theta^T I = \qquad \nabla_A \operatorname{tr} AB = B^T$$

$$X^T X \theta + X^T X \theta$$

2b)

$X^{m\times n}$

$\theta^{n\times 1}$

$y^{n\times 1}$

$W^{m\times m}$ ~~n×n~~

$$\nabla_\theta \cancel{\operatorname{tr}} (X\theta - \bar{y})^T W (X\theta - \bar{y}) =$$

$$= (\theta^T X^T W - \bar{y}^T W)(X\theta - \bar{y})$$

$$= \theta^T X^T W X \theta - \theta^T X^T W \bar{y} - \bar{y}^T W X \theta + \bar{y}^T W \bar{y}$$

$(X\theta)^{m\times 1}$

$$= \nabla_\theta \operatorname{tr} (\theta^T X^T W X \theta - \theta^T X^T W \bar{y} - \bar{y}^T W X \theta + \bar{y}^T W \bar{y})$$

$(X\theta^T)^{1\times m}$

$((X\theta)^T W)^{1\times m}$

$$= \nabla_\theta \underbrace{\operatorname{tr} \theta^T X^T W X \theta} - 2 \underbrace{\operatorname{tr} \bar{y}^T W X \theta}$$

$$- 2 X^T W \bar{y}$$

$$\operatorname{tr} \theta^T X^T W \bar{y} = \operatorname{tr} \bar{y}^T W X \theta$$

b/c $(\theta^T X^T W \bar{y})^T = \bar{y}^T W^T X \theta$

$W = W^T$ (W = diagonal and quadratic

$A^T = \theta$

$B = X^T W X$

$A = \theta^T$

$C = I$

$$\Rightarrow X^T W^T X \theta + X^T W X \theta - 2 X^T W \bar{y}$$

$$= 2 X^T W X \theta - 2 X^T W \bar{y} \qquad \nabla_A \operatorname{tr} AB = B^T$$

$$\Rightarrow X^T W X \theta = X^T W \bar{y} \Rightarrow \boxed{\theta = (X^T W X)^{-1} X^T W \bar{y}}$$

2d

$$L(\theta) = L(\theta; X, \vec{y}) = p(\vec{y} \mid X; \theta)$$

$$L(\theta) = \prod_{i=1}^{m} p(y^{(i)} \mid x^{(i)}; \theta, \sigma^{(i)})$$

$$= \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi}\,\sigma^{i}} \exp\left(-\frac{(y^{(i)} - \theta^T x^{(i)})^2}{2\sigma^{i2}}\right)$$

$\Rightarrow$

$$\ell(\theta) = \log L(\theta)$$

$$= \log \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi}\,\sigma^{(i)}} \exp\left(-\frac{(y^{(i)} - \theta^T x^{(i)})^2}{2\sigma^{(i)2}}\right)$$

$$= \sum_{i=1}^{m} \left(\log \frac{1}{\sqrt{2\pi}}\sigma^{(i)} \cancel{\log} - \frac{(y^{(i)} - \theta^T x^{(i)})^2}{2\sigma^{(i)2}}\right)$$

max likelihood

$$= \cancel{m \log} \sum_{i=1}^{m} \log \frac{1}{\sqrt{2\pi}\,\sigma^{(i)}} - \sum_{i=1}^{m} \frac{(y^{(i)} - \theta^T x^{(i)})^2}{2\sigma^{(i)2}}$$

Const $\Rightarrow$ doesn't matter to minimize

$$\Rightarrow \min \frac{1}{2} \sum_{i=1}^{m} \frac{(y^{(i)} - \theta^T x^{(i)})^2}{\cancel{2}\sigma^{(i)2}}$$

Weighted lin. reg (from lecture notes) $\sum_{i=1}^{m} w^{(i)} (y^{(i)} - \theta^T x^{(i)})^2$

$$\Rightarrow w_i = \frac{1}{\cancel{2}\sigma^{(i)2}} \quad \forall i \in \{1 \dots m\}$$