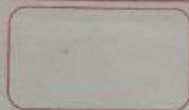


# ML HW 1 T. 1



ЧИСЛО

МІСЯЦЬ

20

РІК

ПН

ВТ

СР

ЧТ

ПТ

СБ

НД

$$1. a. J(\theta) = \frac{1}{2m} \sum_{i=1}^m w^{(i)} (\theta^T x^{(i)} - y^{(i)})^2$$

$$(1) \underset{\theta}{\text{ArgMax}} J(\theta) = \underset{\theta}{\text{ArgMax}} 2m \cdot J(\theta)$$

$\vec{y} = [y^{(1)}, \dots, y^{(m)}]$ ;  $X \equiv m \times k$  matrix,  
where  $k$  - number of features

$\vec{x}^{(i)} = [x_1, \dots, x_k]$  - feature vector for  
 $i$ -th training example.

$$(2) \text{ By definition: } \sum_{i=1}^m (\theta^T x^{(i)} - y^{(i)}) = \|\theta^T X - \vec{y}\| =$$

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

(3) For weighted linear regression  
the cost of incorrectly scoring  
 $i$ -th sample is not equal to  
loss cost of  $j$ -th sample, where  
 $i \neq j$ , in general.

For each sample the importance  
weight for predicting it correctly

is given:  $w_i$ . Let  $W$  be  ~~$n \times n$~~   $m \times m$  diagonal matrix, where  $i$ -th diagonal entry is  $w_i$

$$(4) \quad \sum_{i=1}^m w_i (\theta^T x_i - y_i)^2 \equiv (\underbrace{\theta^T x_1 \dots \theta^T x_m}_{X^T \theta} - \underbrace{y_1 \dots y_m}_{\vec{y}})^T W (\underbrace{\theta^T x_1 \dots \theta^T x_m}_{X^T \theta} - \underbrace{y_1 \dots y_m}_{\vec{y}})$$



2.2

ПН

ВТ

СР

ЧТ

ПТ

СБ

НД

Poisson Distribution

 $X \sim \text{Poisson}(\theta)$ , where  $E[X] = \theta$ 

$$P(X|\theta) = \frac{\theta^x}{x!} e^{-\theta}, x=0, 1, \dots$$

$$= \frac{1}{x!} \exp((\ln \theta)x - \theta) =$$

$$= h(x) \exp(\eta(\theta)T(x) - B(\theta))$$

$$h(x) = \frac{1}{x!}; \quad \eta(\theta) = \ln \theta \quad T(x) = x$$

$$e^{\eta(\theta)} = \theta; \quad \theta = e^{\eta}$$



2.  $g(\eta) = E[T(y) | \eta]$  is called canonical response function. Its inverse,  $g^{-1}$ , is called the canonical link function.

$$g(\eta) = E[T(y) | \eta] = \lambda = e^\eta = [\eta = \ln(\lambda)] = e^{\theta^T x}$$

by the third assumption,

$$P(y^{(i)} | x^{(i)}; \theta) = \frac{1}{y^{(i)}!} \exp(\theta^T x^{(i)} y^{(i)} - e^{\theta^T x^{(i)}})$$

$$J(\theta) = \log P(y^{(i)} | x^{(i)}; \theta) = \theta^T x^{(i)} y^{(i)} - e^{\theta^T x^{(i)}} - \ln(y^{(i)}!)$$

$$\frac{\partial J(\theta)}{\partial \theta_j} = x_j^{(i)} y^{(i)} - x_j^{(i)} \exp(\theta^T x^{(i)}) = x_j^{(i)} (y^{(i)} - \exp(\theta^T x^{(i)}))$$

$$\Delta \theta_j = -\alpha \frac{\partial J}{\partial \theta_j}, \text{ where } \alpha \text{ is a learning rate}$$

$$\theta_j := \theta_j + \Delta \theta_j$$



$$3.a \quad J(\theta) = -\frac{1}{m} \ell(\theta) = -\frac{1}{m} \sum_{i=1}^m w^{(i)} \left[ y^{(i)} \log h_{\theta}(x^{(i)}) + (1-y^{(i)}) \log(1-h_{\theta}(x^{(i)})) \right]$$

where  $w^{(i)} = \exp(-(x^{(i)} - x)^T(x^{(i)} - x)/2\tau^2)$ ,

$$h_{\theta}(x^{(i)}) = 1 / (1 + \exp(-\theta^T x^{(i)}))$$

$$3.b. \quad \theta_j = \theta_j + \lambda (w^{(i)} (y^{(i)} - h_{\theta}(x^{(i)})) x_j^{(i)})$$

3.c. There is no analytical solution.



4. In our model we have 3 features

$$y = \beta_0 x_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

where  $x_0$  is equal to 1.

The model was trained with  $x_3$   
~~was~~ multiplied by  $(1/1.60934)$ .

To correct the output of the model  
 $y$  must be changed to  $\hat{y}$ , where

$$\hat{y} = y - \beta_3 (1/1.60934) x_3 + \beta_3 \cdot x_3$$