ML HW 1 T. 1

ЧИСЛО МІСЯЦЬ 20 РІК

ПН ВТ СР ЧТ ПТ СБ НД

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

(1) $\underset{\theta}{\mathrm{ArgMax}}\, J(\theta) = \underset{\theta}{\mathrm{ArgMax}}\, 2m\, J(\theta)$

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

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

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

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

(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

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

(4) $\sum_{i=1}^{m} w_i (\cancel{\theta^T x_i} - y_i) \equiv (\cancel{\theta^T X} - \vec{y})^T W (\cancel{\theta^T X} - \vec{y})$

$\cancel{\theta}^T x_i \qquad X\theta \qquad X\theta$

2. a

Poisson Distribution

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

$$p(x \mid \theta) = \frac{\theta^x}{x!} e^{-\theta}, \quad x = 0, 1, \ldots$$

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

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

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

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

2.b $g(\eta) = E[T(y); \eta]$ 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.

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

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

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

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

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

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

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. $\theta_j = \theta_j + \alpha(w^{(i)}(y^{(i)} - h_\theta(x^{(i)}))x_j^{(i)}$

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

$$\hat{y} = y - \beta_3 (1/1.60934) X_3 + \beta_3 \cdot X_3$$