Homework 1 Ostap Kharysh

Supervised learning

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

$w^{(i)}$ - weight of each element

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

$$r^2 = [r_1 \ r_2 \ldots r_m] \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_m \end{bmatrix}$$

r - should be a vector $1 \times m$, so $W = \begin{bmatrix} w_1 & \cdots & & 0 \\ 0 & w_2 & \cdot & 0 \\ 0 & & \cdot & 0 \\ & 0 & & w_m \end{bmatrix}$

$$J(\theta) = [r_1, r_2 \ldots r_m] \cdot \begin{bmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & \cdots & 0 \\ 0 & \cdots & w_3 & 0 \\ 0 & \cdots & 0 & w_m \end{bmatrix} \cdot \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_m \end{bmatrix}$$

$J = r^T W r$

$x_i \in R^n \times \theta \in R^n$ where $X \cdot \theta_{(n \times 1)} = X\theta$

~~y = mat~~ y = vector $(m \times 1)$

$r = X\theta - \vec{y}$

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

W - is a diagonal matrix of weights, it has weight of each element of a sample situated successively.

2. Poison distribution

$$P(y;\lambda) = \frac{e^{-\lambda}\lambda^y}{y!}$$

a) λ - is an average rate of hits per minute

Let's assume there is a Lifetime variable L which describes the time to wait for a first hit.

$$P(L>t) = P(\text{at period } t - \text{no hits}) = $$
$$= \frac{e^{-(\lambda t)}\cdot(\lambda t)^0}{0!} = e^{-\lambda t}$$

$$P(L \leq t) = 1 - e^{-\lambda t}$$

$$\left(1-e^{-\lambda t}\right)' = \lambda e^{-\lambda t} \text{ — density function}$$

$$f(t) = \begin{cases} \lambda e^{-\lambda t} & \text{for } t \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

RV that has DF like above is exponentialy distributed.

$\eta = \ln(\lambda)$

$d(\eta) = \lambda$

$T(y) = y$

$b(y) = \frac{1}{y!}$

Let's change λ^y to e^x, where x is unknown

$\ln(\lambda^y) = \ln(e^x)$

$y\ln(\lambda) = x\ln(e)$

$y\ln(\lambda) = x$

$$\boxed{x = y\cdot\ln(\lambda)}$$

$$P(y;\lambda) = \frac{e^{-\lambda}\lambda^y}{y!} = \frac{e^{-\lambda+y\cdot\ln(\lambda)}}{y!} = \frac{e^{y\ln(\lambda)-\lambda}}{y!}$$

$$P(y;\eta) = b(y)\cdot e^{\eta^T T(y) - d(\eta)} \Rightarrow \eta = \ln(\lambda)$$

single exponential family

$d(\eta) = \lambda$; $b(y) = \frac{1}{y!}$; $T(y) = y$,

b) link function for distribution is $\ln(\mu)$

$$f^{-1}(x) = \ln(x)$$

$$f(x) = e^{x}$$

So the cannonical response function =

$= E[y; \lambda] = \ln(\lambda)$, assumed that $\mu = \lambda$.

$* \ \eta^{T} = \log(\lambda)$ (from a)) ; $e^{\eta^{T}} = e^{\log \lambda} = e^{n^{T}} = \lambda$.

So Cannonical response function $g(\mu) = g(\lambda) = e^{\eta} = e^{\theta^{T} X}$

c) $\ell(\theta) = \log P(y^{(i)} / x^{(i)}; \theta) =$

$= \log(b(y) \exp(\eta(T(y) - a(\eta)))$ – hence single exponatial family

Assume $\eta = \theta^{T} x$

$$\ell(\theta) = \log\left(\frac{1}{y!} \cdot e^{\theta^{T} x (y - e^{n})}\right)$$

$$\ell(\theta) = \log\left(\frac{1}{y!}\right) + \log\left(e^{\theta^{T} x (y - e^{n})}\right) =$$

$$= \log\left(\frac{1}{y!}\right) + \theta^{T} x (y - e^{n})$$

$$\frac{\partial}{\partial \theta_j} \ell(\theta) = \frac{\partial}{\partial \theta_j}\left(\log\left(\frac{1}{y!}\right) + \theta^{T} x (y - e^{n})\right) =$$

$$= \cancel{\frac{\partial}{\partial \theta_j} \log\left(\frac{1}{y!}\right)} + \frac{\partial}{\partial \theta_j}\left(\theta_j^{T} x (y - e^{n})\right) =$$

$$= \frac{\partial \theta_j^{T} x}{\partial \theta_j} \cdot (y - e^{n}) + \cancel{\frac{\partial (y - e^{n})}{\partial \theta_j} \cdot (\theta_j^{T} x)} =$$

$$= x(y - e^{n}) + 0 = x(y - e^{n})$$

$\theta_j = \theta_j + \alpha (x_j (y - e^{n}))$ – update of weights