Στατιστική Μοντελοποίηση και Αναγνώριση Προτύπων

2ο Φυλλάδιο Ασκήσεων

ΘΕΜΑ 1

α) Αρχικά υπολογίζω την παράγωγο $\dfrac{\partial h_\theta(x)}{\partial \theta_j}$

$$\frac{\partial h_\theta(x^{(i)})}{\partial \theta_j} = -\frac{1}{(1+e^{-\theta^T x^{(i)}})^2} e^{-\theta^T x} \cdot \frac{\partial}{\partial \theta_j}(-\theta^T x^{(i)}) = \frac{1}{(1+e^{-\theta^T x^{(i)}})^2} e^{-\theta^T x} \cdot \frac{\partial}{\partial \theta_j}(\theta^T x^{(i)})$$

Υπολογισμός της $\dfrac{\partial}{\partial \theta_j}(\theta^T x^{(i)}) = \dfrac{\partial}{\partial \theta_j} [\theta_1, \theta_2, \dots, \theta_n] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$

$$= \frac{\partial}{\partial \theta_j}\theta_1 x_1 + \frac{\partial}{\partial \theta_j}\theta_2 x_2 + \dots + \frac{\partial}{\partial \theta_j}\theta_j x_j + \dots + \frac{\partial}{\partial \theta_j}\theta_n x_n$$

$$= x_j$$

Άρα, $\dfrac{\partial}{\partial \theta_j} h_\theta(x^{(i)}) = \dfrac{e^{-\theta^T x^{(i)}}}{(1+e^{-\theta^T x^{(i)}})^2} \cdot x_j$

$$= \frac{e^{-\theta^T x^{(i)}} + 1 - 1}{1+e^{-\theta^T x^{(i)}}} \cdot \frac{1}{1+e^{-\theta^T x^{(i)}}} \cdot x_j = (1-h_\theta(x^{(i)}))\, h_\theta(x^{(i)})\, x_j$$

$$\Rightarrow \frac{\partial h_\theta(x^{(i)})}{\partial \theta_j} = (1-h_\theta(x^{(i)}))\, h_\theta(x^{(i)}) \cdot x_j$$

Η ζητούμενη κλίση είναι:

$$\frac{\partial J(\theta)}{\partial \theta_j} = \frac{1}{m}\sum_{i=1}^{m}\left(-y^{(i)}\ln\left(h_\theta(x^{(i)})\right) - (1-y^{(i)})\ln\left(1-h_\theta(x^{(i)})\right)\right)$$

$$= \frac{1}{m}\sum_{i=1}^{m}\left(-y^{(i)}\cdot\frac{\partial h_\theta(x^{(i)})/\partial\theta_j}{h_\theta(x^{(i)})} - (1-y^{(i)})\frac{\partial h_\theta(x^{(i)})/\partial\theta_j}{1-h_\theta(x^{(i)})}\right)$$

$$= \frac{1}{m}\sum_{i=1}^{m}\left(-y^{(i)}\cdot\frac{(1-h_\theta(x^{(i)}))\cancel{h_\theta(x^{(i)})}\cdot x_j}{\cancel{h_\theta(x^{(i)})}} - (1-y^{(i)})\frac{\cancel{(1-h_\theta(x^{(i)}))}h_\theta(x^{(i)})\,x_j}{\cancel{1-h_\theta(x^{(i)})}}\right)$$

$$= \frac{1}{m}\sum_{i=1}^{m}\left(h_\theta(x^{(i)}) - y^{(i)}\right)\cdot x_j^{(i)}$$

Όπερ έδει δείξαι

ΘΕΜΑ 2

α) Ο πρώτος όρος της ζητούμενης παραγώγου έχει αποδειχθεί στο προηγούμενο θέμα. (Θέμα 1 - α)

Για τον δεύτερο όρο:

$$\frac{\partial}{\partial \theta_j}\left(\frac{\lambda}{2m}\sum_{j=1}^{n}\theta_j^2\right) = \frac{\partial}{\partial\theta_j}\left[\frac{\lambda}{2m}\left(\theta_1^2+\theta_2^2+\theta_3^2+\dots+\theta_n^2\right)\right]$$

$$= \frac{\lambda}{2m}\left[\frac{\partial}{\partial\theta_j}\theta_1^2 + \frac{\partial}{\partial\theta_j}\theta_2^2 + \dots + \frac{\partial}{\partial\theta_j}\theta_j^2 + \dots + \frac{\partial}{\partial\theta_j}\theta_n^2\right] = \frac{\lambda}{2m}\left(0+0\dots+2\theta_j+0\right)$$

$$= \frac{\lambda}{m}\theta_j$$

Άρα

$$\boxed{\frac{\partial J(\theta)}{\partial \theta_j} = \frac{1}{m}\sum_{i=1}^{m}\left(h_\theta(x^{(i)}) - y^{(i)}\right)x_j^{(i)} + \frac{\lambda}{m}\theta_j}$$

ΤΗΛ 311 - ΣΜΑΠ

2η εργασία

▶ Θέμα 3ο:

• $\mathcal{D}=\{x_1,\dots,x_n\}$, $p(x|\lambda)=\dfrac{\lambda^x e^{-\lambda}}{x!}$, $x=0,1,2,\dots$, $\lambda>0$

α) $\hat{\lambda}_{ML}=;$ $\left(\hat{\theta}_{ML}=\max\limits_{\theta} p(\mathcal{D}|\theta)\right) \rightarrow$ εδώ όπου $\theta\equiv\lambda$.

$\cdot\, p(\mathcal{D}|\lambda)=p(x_1,\dots,x_n|\theta)=p(x_1|\lambda)\cdot\ldots\cdot p(x_n|\lambda)=$

$$=\frac{\lambda^{x_1}e^{-\lambda}}{x_1!}\cdot\ldots\cdot\frac{\lambda^{x_n}e^{-\lambda}}{x_n!}=\prod_{k=1}^{n}\frac{\lambda^{x_k}\cdot e^{-\lambda}}{x_k!}$$

$\cdot\, \hat{\lambda}_{ML}=\operatorname{argmax}\limits_{\lambda} p(\mathcal{D}|\lambda) \Leftrightarrow \hat{\lambda}_{ML}=\operatorname{argmax}\limits_{\lambda} \mathcal{L}(\lambda)$

$$\mathcal{L}(\lambda)=\ln p(\mathcal{D}|\lambda)=\ln\left(\prod_{k=1}^{n}p(x_k|\lambda)\right)=\sum_{k=1}^{n}\ln p(x_k|\lambda) \ ⊛$$

$$p(x_k|\lambda)=\frac{\lambda^{x_k}e^{-\lambda}}{x_k!}\rightarrow \ln p(x_k|\lambda)=\ln\left[\frac{\lambda^{x_k}\cdot e^{-\lambda}}{x_k!}\right]=$$

$$=\ln\left(\lambda^{x_k}\cdot e^{-\lambda}\right)-\ln(x_k!)=$$

$$=\ln\lambda^{x_k}-\ln x_k!-\lambda$$

$\cancel{⊛\ \mathcal{L}(\lambda)=\sum_{k=1}^{n}\ln\lambda^{x_k}-\ln x_k!-\lambda}$

$\cancel{\frac{\partial\mathcal{L}(\lambda)}{\partial\lambda}=0 \Leftrightarrow \frac{\partial\mathcal{L}(\lambda)}{\partial\lambda}=}$

$$⊛\ \mathcal{L}(\lambda)=\sum_{k=1}^{n}\ln\lambda^{x_k}-\ln x_k!-\lambda=\sum_{k=1}^{n}\ln\lambda^{x_k}-\sum_{k=1}^{n}\ln x_k!-\sum_{k=1}^{n}\lambda=$$

$$=\ln\lambda\sum_{k=1}^{n}x_k-\sum_{k=1}^{n}\ln x_k!-\lambda\sum_{k=1}^{n}1=$$

$$=\ln\lambda\sum_{k=1}^{n}x_k-\sum_{k=1}^{n}\ln x_k!-\lambda N$$

→

$$\frac{\partial \mathcal{L}(\lambda)}{\partial \lambda}=0 \Rightarrow \frac{\partial \ln\lambda}{\partial \lambda}\cdot\sum_{k=1}^{N} x_k - \frac{\partial}{\partial\lambda}\left(\cancel{\sum_{k=1}^{N}\ln x_k!}\right)^{0} - \frac{\partial(N\lambda)}{\partial\lambda} = 0 \Rightarrow$$

$$\Rightarrow \frac{1}{\lambda}\cdot\sum_{k=1}^{N} x_k - N = 0 \Rightarrow \frac{1}{\lambda}\sum_{k=1}^{N} x_k = N \Rightarrow$$

$$\Rightarrow \frac{1}{N}\sum_{k=1}^{N} x_k = \lambda \Rightarrow \boxed{\hat{\lambda}_{ML} = \bar{x}}$$ mean

b) $\hat{\lambda}_{MAP} = ;$

- $p(\lambda|D) = \frac{p(\lambda, D)}{p(D)} = \frac{p(D|\lambda)\cdot p(\lambda)}{p(D)}$
- $\hat{\lambda}_{MAP} = \arg\max_{\lambda} p(\lambda|D) = \arg\max_{\lambda} p(D|\lambda)\cdot p(\lambda) \overset{a)}{=} \arg\max_{\lambda}\left[\left(\prod_{k=1}^{N}\frac{\lambda^{x_k}e^{-\lambda}}{x_k!}\right)\left(e^{-\lambda}\right)\right]$
- $\hat{\lambda}_{MAP} = \arg\max_{\lambda} p(\lambda|D) \Leftrightarrow \hat{\lambda}_{MAP} = \arg\max_{\lambda} \mathcal{L}(\lambda)$

  - $\mathcal{L}(\lambda) = \ln p(\lambda|D) = \ln p(D|\lambda) + \ln p(\lambda) \overset{a}{=}$

  $= \ln\lambda\cdot\sum_{k=1}^{N} x_k - \sum_{k=1}^{N}\ln(x_k!) - \lambda N + \ln(e^{-\lambda}) = \ln\lambda\cdot\sum_{k=1}^{N} x_k - \sum_{k=1}^{N}\ln(x_k!) - \lambda N - \lambda =$

  $= \ln\lambda\cdot\sum_{k=1}^{N} x_k - \sum_{k=1}^{N}\ln(x_k!) - \lambda(N+1)$

  - $\frac{\partial\mathcal{L}(\lambda)}{\partial\lambda} = 0 \Rightarrow \frac{\partial}{\partial\lambda}\left(\ln\lambda\cdot\sum_{k=1}^{N} x_k\right) - \frac{\partial}{\partial\lambda}\left(\cancel{\sum_{k=1}^{N}\ln(x_k!)}\right)^{0} - \frac{\partial}{\partial\lambda}\left(\lambda(N+1)\right) = 0 \Rightarrow$

  $\Rightarrow \frac{1}{\lambda}\sum_{k=1}^{N} x_k - (N+1) = 0 \Rightarrow \frac{1}{\lambda}\sum_{k=1}^{N} x_k = N+1 \Rightarrow \frac{1}{N+1}\sum_{k=1}^{N} x_k = \lambda \Rightarrow$

  $\Rightarrow \boxed{\hat{\lambda}_{MAP} = \frac{1}{N+1}\sum_{k=1}^{N} x_k}$ normilized mean

ΘΕΜΑ 5

α)

$$J(y^{(i)}, \hat{y}^{(i)}; w, b) = -y^{(i)} \ln(\hat{y}^{(i)}) - (1-y^{(i)}) \ln(1-\hat{y}^{(i)})$$

$$z^{(i)} = x^{(i)} w + b \Rightarrow \hat{y}^{(i)} = f(z^{(i)})$$

$$1 - \frac{1}{1+e^{-z}} = \frac{e^{-z}}{1+e^{-z}}$$

$$J(y^{(i)}, \hat{y}^{(i)}; w, b) = -y^{(i)} \ln\left(\left(1+e^{-z^{(i)}}\right)^{-1}\right) - (1-y^{(i)}) \ln\left(1-\left(1+e^{-z^{(i)}}\right)^{-1}\right)$$

$$= y^{(i)} \cdot \ln(1+e^{-z^{(i)}}) - (1-y^{(i)}) \cdot \ln\left(\frac{e^{-z^{(i)}}}{1+e^{-z^{(i)}}}\right)$$

$$= y \ln(1+e^{-z}) - \ln\left(\frac{e^{-z}}{1+e^{-z}}\right) + y \ln\left(\frac{e^{-z}}{1+e^{-z}}\right)$$

$$= y \ln(1+e^{-z}) - \left(-z - \ln(1+e^{-z})\right) + y\left(-z - \ln(1+e^{-z})\right)$$

$$= \cancel{y \ln(1+e^{-z})} + z + \ln(1+e^{-z}) - y \cdot z - \cancel{y \ln(1+e^{-z})}$$

$$= z^{(i)} - y^{(i)} \cdot z^{(i)} + \ln(1+e^{-z^{(i)}})$$

β) $$\frac{\partial J}{\partial z^{(i)}} = \frac{\partial}{\partial z^{(i)}} \left(\frac{1}{B} \sum_j \left(z^{(j)} - z^{(j)} y^{(j)} + \ln(1+e^{-z^{(j)}}) \right) \right)$$

$$= \frac{1}{B}\left(1 - y^{(i)} - \frac{1}{1+e^{-z^{(i)}}} e^{-z^{(i)}}\right) = \frac{1}{B} \cdot \left(1 - y^{(i)} - \left(1 - \frac{1}{1+e^{-z^{(i)}}}\right)\right) = \frac{1}{B} \left(-y^{(i)} \ \hat{y}^{(i)}\right)$$

γ). $\frac{\partial J}{\partial W} = \frac{\partial J}{\partial z^{(i)}} \frac{\partial z^{(i)}}{\partial w}$

$$\frac{\partial J}{\partial z^{(i)}} = \frac{1}{B} \cdot \left(-y^{(i)} + \hat{y}^{(i)}\right)$$

$$\frac{\partial z^{(i)}}{\partial w} = \frac{\partial}{\partial w} (x^{(i)} w + b) = x^{T(i)}$$

Άρα $\frac{\partial J}{\partial W} = \frac{1}{B} \cdot \left(-y^{(i)} + \hat{y}^{(i)}\right) \cdot x^{T(i)}$

. $\frac{\partial J}{\partial b} = \frac{\partial J}{\partial z^{(i)}} \cdot \frac{\partial z^{(i)}}{\partial b} = \frac{1}{B} \left(-y^{(i)} + \hat{y}^{(i)}\right)$.

$$\frac{\partial z^{(i)}}{\partial b} = \frac{\partial}{\partial b} \left(x^{(i)} w + b\right) = 0 + 1 = 1$$