Θέμα 1: Λογιστική Παλινδρόμηση: Αναλυτική εύρεση κλίσης (Gradient).

a) m δεδομένα $\{(x^{(1)}, y^{(1)}), \dots, (x^{(m)}, y^{(m)})\}$

$x^{(i)} \in \mathbb{R}^{n\times 1}$, $y^{(i)} \in \{0,1\}$ → κλάση κάθε δείγματος

↳ διάνυσμα χαρακτηριστικών.

Συνάρτηση λογιστικής παλινδρόμησης:

$$h_\theta(x) = f(\theta^T x), \quad \theta = [\theta_1, \theta_2, \dots, \theta_n]^T$$

$$= \frac{1}{1+e^{-\theta^T x}} \quad \text{και } x = [x_1, x_2, \dots x_n]^T.$$

Υπολογίζω την παράγωγο: $\frac{\partial h_\theta(x)}{\partial \theta_j}$

$$\boxed{\frac{\partial}{\partial \theta_j}(h_\theta(x))} = \frac{\partial}{\partial \theta_j}\left(\frac{1}{1+e^{-\theta^T x}}\right)$$

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

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

$$= \frac{e^{-\theta^T x}}{(1+e^{-\theta^T x})^2} \cdot \frac{\partial}{\partial \theta_j}\left(\sum_{k=1}^{n} \theta_k \cdot x_k\right) = \boxed{\frac{x_j e^{-\theta^T x}}{(1+e^{-\theta^T x})^2}}$$

Υπολογίζω έπειτα την σχέση: $1-h_\theta(x)$.

$$\boxed{1-h_\theta(x)} = 1 - \frac{1}{1+e^{-\theta^T x}} = \boxed{\frac{e^{-\theta^T x}}{1+e^{-\theta^T x}}}$$

Ζητείται νδο:

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

$$\Rightarrow \boxed{\frac{\partial}{\partial\theta_j}(J(\theta))}=\frac{1}{m}\sum_{i=1}^{m}-y^{(i)}\cdot\frac{1}{h_\theta(x^{(i)})}\frac{\partial}{\partial\theta_j}(h_\theta(x^{(i)}))$$

$$+(1-y^{(i)})\cdot\frac{1}{1-h_\theta(x^{(i)})}\frac{\partial}{\partial\theta_j}(h_\theta(x^{(i)})).$$

Χρήση των 2 σχέσεων

που βρέθηκαν νωρίτερα

$$\frac{1}{m}\sum_{i=1}^{m}-y^{(i)}(1+e^{-\theta^T x})^{(i)}\left(\frac{x_j^{(i)}\cdot e^{-\theta^T x^{(i)}}}{(1+e^{-\theta^T x})^2}\right)$$

$$+(1-y^{(i)})\frac{1+e^{-\theta^T x^{(i)}}}{e^{-\theta^T x}}\cdot\left(\frac{x_j^{(i)}\cdot e^{-\theta^T x^{(i)}}}{(1+e^{-\theta^T x})^2}\right)$$

$$=\frac{1}{m}\sum_{i=1}^{m}-y^{(i)}x_j^{(i)}\cdot\frac{e^{-\theta^T x^{(i)}}}{(1+e^{-\theta^T\cdot x^{(i)}})}+\frac{(1-y^{(i)})\cdot x_j^{(i)}}{(1+e^{-\theta^T\cdot x^{(i)}})}$$

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

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

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

Scanned with CamScanner

Θέμα 2: Λογιστική Παλινδρόμηση με Ομαδοποίηση.

α) Έχω $\theta_j, x_j^{(i)}$ να είναι η j-οστή συνιστώσα των διανυσμάτων $\theta = [\theta_1, \theta_2, \dots, \theta_n]^T$ και $x^{(i)} = [x_1^{(i)}, x_2^{(i)}, \dots, x_n^{(i)}]^T$.

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{K} -y^{(i)} \ln(h_\theta(x^{(i)}) - (1-y^{(i)}) \cdot \ln(1-h_\theta(x^{(i)}))$$

$$+ \frac{\lambda}{2m}\sum_{j=1}^{K}\theta_j^2$$

$$\boxed{\frac{\partial J(\theta)}{\partial \theta_j}} = \frac{1}{m}\frac{\partial}{\partial \theta_j}\Big(\underbrace{\sum -y^{(i)} \ln(h_\theta(x^{(i)})) - (1-y^{(i)}) \cdot \ln(1-h_\theta(x^{(i)}))}_{}\Big)$$

↳ χρήση του θέματος.

$$+ \frac{\partial}{\partial \theta_j}\left(\frac{\lambda}{2m}\sum_{j=1}^{K}\theta_j^2\right)$$

$$= \frac{1}{m}\sum_{K=1}^{m} x_j^{(i)}(h_\theta(x^{(i)}) - y^{(i)}) + \frac{\lambda}{2m}\sum_{j=1}^{K}\frac{\partial}{\partial \theta_j}\theta_j^2$$

$$= \boxed{\frac{1}{m}\sum_{K=1}^{m} x_j^{(i)} \cdot (h_\theta(x^{(i)}) - y^{(i)}) + \frac{\lambda}{m}\theta_j}$$

Θέμα 3: Εκτίμηση Παραμέτρων (Maximum Likelihood)

n δείγματα, $D = \{x_1 \dots x_n\}$, ανεξάρτητα από κατανομή Poisson με παράμετρο λ:

$$p(x|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0,1,2\dots, \quad \lambda > 0.$$

Να εκτίμηση το λ με M.L.E

$$L_D(\lambda) = P(D|\lambda) = \prod_{k=1}^{n} p(x_k|\lambda)$$

$$\lambda_{MLE} = \arg\max_{\lambda} P(D|\lambda)$$

$$l_D(\lambda) = \ln L_D(\lambda) = \sum_{k=1}^{n} \ln p(x_k|\lambda)$$

Πρέπει να ισχύει: $\frac{\partial}{\partial \lambda}(l_D(\lambda)) = 0$

$$\Rightarrow \frac{\partial}{\partial \lambda}\left(\sum_{k=1}^{n} \ln p(x_k|\lambda)\right) = 0$$

$$\Rightarrow \sum_{k=1}^{n} \frac{\partial}{\partial \lambda} \ln \left(\frac{\lambda^{x_k} e^{-\lambda}}{x!}\right) = 0$$

$$\Rightarrow \sum_{k=1}^{n} \frac{x!}{\lambda^{x_k} e^{-\lambda}} \cdot \frac{\partial}{\partial \lambda}\left(\frac{\lambda^{x_k} e^{-\lambda}}{x!}\right) = 0$$

$$\Rightarrow \sum_{k=1}^{n} \frac{\cancel{x!}}{\lambda^{x_k} e^{-\lambda}} \cdot \frac{x_k \lambda^{x_k - 1} e^{-\lambda} - \lambda^{x_k} e^{-\lambda}}{\cancel{x!}} = 0$$

$$\Rightarrow \sum_{k=1}^{n} \frac{\lambda^{x_k} e^{-\lambda}\left(\frac{x_k}{\lambda} - 1\right)}{\lambda^{x_k} e^{-\lambda}} = 0$$

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

$$\Rightarrow \boxed{\lambda_{MLE} = \frac{1}{n} \cdot \sum_{k=1}^{n} x_k}$$

Θέμα 4: Support Vector Machines
(Αναλυτική βελτιστοποίηση με KKT)

Lagrangian:

$$L(w,w_0,\lambda)=\frac{1}{2}(w_1^2+w_2^2)-\sum \lambda_i\left[y_i(w^T x_i+w_0)-1\right]$$

$$=\frac{w_1^2+w_2^2}{2}-\lambda_1(-(-10w_1-10w_2+w_0)-1)-\lambda_2(-(-7w_1-7w_2+w_0)-1)$$

Έχω από το $\frac{\partial}{\partial w}L(w,w_0,\lambda)$:

K.K.T: $\frac{\partial}{\partial w_1}L(w,w_0,\lambda)=0 \Rightarrow w_1=10\lambda_1-7\lambda_2$

$\frac{\partial}{\partial w_2}L(w,w_0,\lambda) \Rightarrow w_2=10\lambda_1-7\lambda_2$.

$\Rightarrow$ Άρα $\boxed{w_1=w_2}$ (1)

Αφού έχω 2 support vectors:

$\lambda_1\left(-\left((w_1 \cdot w_2)\binom{-10}{-10}+w_0\right)-1\right)=0$

$\lambda_2\left((w_1 \cdot w_2)\binom{-7}{-7}+w_0-1\right)=0$

$\boxed{10w_1+10w_2-w_0-1=0}$ (2) ΚΑΙ

$\boxed{-7w_1-7w_2+w_0-1=0}$ (3)

(1) και (2) $\Rightarrow 20w_1-w_0-1=0$
(1) και (3) $\Rightarrow -14w_1+w_0-1=0$ } τις προσθέτω

$\Rightarrow 6w_1-2=0 \Rightarrow \boxed{w_1=\frac{1}{3}}$

~~[illegible]~~

από την (1) $\Rightarrow \boxed{w_2 = \frac{1}{3}}$

Επομένως, $\boxed{w_0 = 2 - 4w_1 - 1 = \frac{17}{3}}$

Τελικά $w^T x + w_0 = 0 \Rightarrow \left(\frac{1}{3} \ \frac{1}{3}\right)\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \frac{17}{3} = 0$

$\Rightarrow \frac{x_1}{3} + \frac{x_2}{3} + \frac{17}{3} = 0 \Rightarrow \boxed{x_1 = -x_2 - 17}$

Scanned with CamScanner

Θέμα 5: Υλοποίηση ενός απλού νευρωνικού δικτύου.

Μέρος Α:

Αρχικά έχω τις σχέσεις:

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

και

$$\boxed{J(Y,\hat{Y};W,b) = \frac{1}{B}\sum_i\left(-y^{(i)}\ln\hat{y}^{(i)} - (1-y^{(i)})\ln(1-\hat{y}^{(i)})\right)} \quad (2)$$

α) Θέτω στην σχέση (2):

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

$$\Rightarrow \boxed{J(Y,\hat{Y};W,b)} = \frac{1}{B}\sum_i -y^{(i)}\ln\frac{1}{1+e^{-z^{(i)}}} - (1-y^{(i)})\ln\frac{e^{-z^{(i)}}}{1+e^{-z^{(i)}}}$$

Άρα

$$= \frac{1}{B}\sum_i -y^{(i)}\left(\cancel{\ln 1}^{0} - \ln(1+e^{-z^{(i)}})\right) - (1-y^{(i)})\left(\ln e^{-z^{(i)}} - \ln(1+e^{-z^{(i)}})\right)$$

$$= \frac{1}{B}\sum_i \cancel{y^{(i)}\ln(1+e^{-z^{(i)}})} - \ln e^{-z^{(i)}} + \ln(1+e^{-z^{(i)}}) + y^{(i)}\ln e^{-z^{(i)}}$$

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

$$= \boxed{\frac{1}{B}\sum_i z^{(i)} - y^{(i)}z^{(i)} + \ln(1+e^{-z^{(i)}})}$$

β) Παραγωγίζω την σχέση (1) αφού την γράψω ως συνάρτηση του $z^{(i)}$.

Σημειωμένα, δεν θέλω το αποτέλεσμα κανονικοποιημένο με το batch size B που έχω. Θα δουλέψω πάνω σε 1 sample.

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

$= -y^{(i)}\ln(f(z^{(i)})) - (1-y^{(i)})\ln(1-f(z^{(i)}))$

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

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

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

Παραγωγίζω ως προς $z^{(i)}$: $\boxed{\frac{\partial J(y^{(i)},\hat{y}^{(i)};w,b)}{\partial z^{(i)}}} =$

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

$= 1 - y^{(i)} + \frac{1}{1+e^{-z^{(i)}}}(-e^{-z^{(i)}}) = \boxed{\hat{y}^{(i)} - y^{(i)}}$

2) Για $\frac{\partial J}{\partial w}$:

$\Rightarrow$ Για sample:

$\boxed{\frac{\partial J(y^{(i)},\hat{y}^{(i)};w,b)}{\partial w}} = \frac{\partial J(y^{(i)},\hat{y}^{(i)};w,b)}{\partial z^{(i)}}\cdot\frac{\partial z^{(i)}}{\partial w}$

$= \boxed{(\hat{y}^{(i)} - y^{(i)})\cdot x^{(i)}}$

για batch B:

$$J(Y,\hat{Y};W,b) = \frac{1}{B}\sum_i J(y^{(i)},\hat{y}^{(i)};W,b)$$

$$\boxed{\frac{\partial J(Y,\hat{Y};W,b)}{\partial W}} = \frac{1}{B}\sum_i \frac{\partial J(y^{(i)},\hat{y}^{(i)};W,b)}{\partial W}$$

$$= \boxed{\frac{1}{B}\sum_i (\hat{y}^{(i)} - y^{(i)})\cdot x^{(i)}}$$

↳ χρήση στον αλγόριθμο gradient descent στον κώδικα

Για $\frac{\partial J}{\partial b}$.

για sample:

$$\boxed{\frac{\partial J(y^{(i)},\hat{y}^{(i)};W,b)}{\partial b}} = \frac{\partial J(y^{(i)},\hat{y}^{(i)};W,b)}{\partial z^{(i)}}\cdot\frac{\partial z^{(i)}}{\partial b}$$

$$= (\hat{y}^{(i)} - y^{(i)})\cdot 1 = \boxed{\hat{y}^{(i)} - y^{(i)}}$$

για batch B:

$$J(Y,\hat{Y};W,b) = \frac{1}{B}\sum_i J(y^{(i)},\hat{y}^{(i)};W,b)$$

$$\boxed{\frac{\partial J(Y,\hat{Y};W,b)}{\partial b}} = \frac{1}{B}\cdot\sum_i \frac{\partial J(y^{(i)},\hat{y}^{(i)};W,b)}{\partial b}$$

$$= \boxed{\frac{1}{B}(\hat{y}^{(i)} - y^{(i)})}$$