Φροντιστήριο

Leniar Regration ∇ Gradient

1) Κόστος: $J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} \big(\overbrace{h_\theta(x^{(i)})}^{h(x^{(i)};\theta)} - y^{(i)} \big)^2$

Ελαχίστων Τετραγώνων

πίνακας $\rightarrow X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix}$

στοιχείο εγγραφής $\rightarrow x_m$
π.χ. εικόνα

αριθμός δειγμάτων → $m \times d$ ← πλήθος features

$$\nabla_\theta J = \begin{bmatrix} \frac{\partial J}{\partial \theta_0} \\ \frac{\partial J}{\partial \theta_1} \\ \vdots \\ \vdots \end{bmatrix} \xleftrightarrow[\text{διάσταση}]{\text{ίδια}} \theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \vdots \\ \theta_d \end{bmatrix} \in \mathbb{R}^d$$

d-άδα

$\theta_i \in \mathbb{R}$

$\theta \leftarrow \theta - n \underbrace{\nabla_\theta J}_{\in \mathbb{R}}$

$$\frac{\partial J}{\partial \theta_i} = \frac{1}{2m} \sum_{i=1}^{m} \frac{\partial}{\partial \theta_j} [h_\theta(x^{(i)}) - y^{(i)}]^2$$

$$= \frac{1}{\cancel{2}m} \sum_{i=1}^{m} \cancel{2}(h_\theta(x^{(i)}) - y^{(i)}) \frac{\partial}{\partial \theta_j} (h_\theta(x^{(i)}) - y^{(i)}) = *$$

$$\Bigg(\underset{1\times 1}{h_\theta(x^{(i)})} = \underset{1\times d}{\theta^T} \underset{d\times 1}{x^{(i)}} = \sum_{k=0}^{d} \theta_k x_k^{(i)} = \theta_j x_j^{(i)} + \sum_{k \neq j} \theta_k x_k^{(i)}$$

$$\Rightarrow \boxed{\frac{\partial h_\theta}{\partial \theta_j} = x_j^{(i)}} \qquad \frac{\partial ax}{\partial x} = a \Bigg)$$

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

$$\theta^{k+1} \leftarrow \theta^k - n \nabla_\theta J$$

$\ln'(x) = \frac{1}{x}$, Μοντέλο $Y = \sigma(AX)$ Σιγμοειδής

$\ln(g(x))' = \frac{1}{g(x)} g'(x)$ $\sigma(z) = \frac{1}{1+e^{-z}}$, $0 \le \sigma \le 1$

2) Logistic Regration

→ Στις διαφάνειες έχει ένα ~~[illegible]~~ descend

Κόστος: $J(\theta) = \sum_{i=1}^{m} [y^{(i)} \ln(h_\theta(x^{(i)})) + (1-y^{(i)}) \ln(1-h_\theta(x^{(i)}))$

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

$$\frac{\partial}{\partial \theta_j}\left(\ln(h_\theta(x^{(i)}))\right) = \frac{1}{h_\theta(x^{(i)})} \frac{\partial h_\theta(x^{(i)})}{\partial \theta_j} = \frac{1}{h_\theta(x^{(i)})}\left[\frac{\partial}{\partial \theta_j} \sigma(\theta^T x^{(i)})\right]$$

$$= \frac{1}{h_\theta(x^{(i)})} \sigma'(\theta^T x^{(i)}) \cancel{\frac{\partial}{\partial \theta_j} \theta^T(x^{(i)})} \to x_j^{(i)} =$$

Το υπολογίσαμε προηγουμένως $\sigma'(z) = \sigma(z)(1-\sigma(z))$

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

$$\frac{\partial}{\partial \theta_j} \ln(1-h_\theta(x^{(i)})) = \frac{1}{1-h_\theta(x^{(i)})} \frac{\partial}{\partial \theta_j}(1-h_\theta(x^{(i)})) =$$

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

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

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

$$= \sum_{i=1}^{m}\left[y^{(i)} - \cancel{y^{(i)}h_\theta(x^{(i)})} - h_\theta(x^{(i)}) + \cancel{y^{(i)}h_\theta(x^{(i)})}\right]x_j^{(i)}$$

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

Εγγραφή n

$y_n = \sigma(a_n)$, $a_n = \theta^T x_n \rightarrow \dfrac{\partial a_n}{\partial \theta_j} = \dfrac{\partial}{\partial \theta_j}\theta^T x_n = x_{nj}$

σιγμοειδής συνάρτηση

Εγγραφή n
Attribute j

$$\frac{\partial y_n}{\partial a_n}\sigma(a_n) = \sigma(a_n)(1-\sigma(a_n))$$

Κόστος για την εγγραφή n $E_n = t_n \log(y_n) + (1-t_n)\log(1-y_n)$

~~πραγματική εκτίμηση~~
πραγματικό label εδώ

$$\frac{\partial E_n}{\partial \theta_j} = \frac{\partial E_n}{\partial y_n}\frac{\partial y_n}{\partial a_n}\frac{\partial a_n}{\partial \theta_j}$$

$$\frac{\partial E_n}{\partial y_n} = t_n\frac{\partial \log(y_n)}{\partial y_n} + (1-t_n)\frac{\partial \log(1-y_n)}{\partial y_n} =$$

$$= t_n \frac{1}{y_n} + (1-t_n)\frac{1}{1-y_n}(-1) = \frac{t_n - y_n}{y_n(1-y_n)}$$

$$\frac{\partial E}{\partial \theta_j} = \left(\frac{t_n - y_n}{\cancel{y_n(1-y_n)}}\right)\cancel{y_n(1-y_n)}\, x_{nj} = (t_n - y_n)x_{nj}$$

$$\forall n \in [1, m]$$

$$\sum_{n=1}^{m} E_n = \sum (t_n - \underbrace{y_n}_{h_\theta(x_n)})\, x_{nj}$$

Φροντιστήριο

2.1 Έστω ότι εφαρμόζουμε την τεχνική cross validation για την εύρεση συγχρόνως δύο παραμέτρων κανονικοποίησης (regularization) λ_1 και λ_2

Σκοπεύουμε να εξετάσουμε 100 διαφορετικές τιμές για την λ_1 και 100 διαφορετικές για την λ_2

Στο cross validation χωρίζουμε το σύνολο εκπαίδευσης σε $k=10$ κομμάτια. Υπολογίστε πόσες φορές πρέπει να τρέξετε τα πειράματα.

Λύση

$$E = \underbrace{CCE}_{} + \overset{100}{\lambda_1} \|W_1\|^2 + \overset{100}{\lambda_2} \|W_2\|^2$$

πείραμα

Cross Entropy Loss $\sum_{k=1}^{K} \sum_{n=1}^{N} t_n \log(y_{nk})$

Άρα 100x100 για τις παραμέτρους

$K = 10$ $\quad |\Lambda_1| \; |\Lambda_2| \; K$

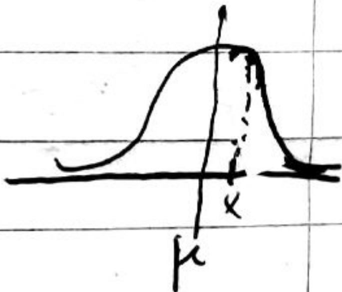
Οπότε στο σύνολο $100 \cdot 100 \cdot 10 = 10^5$

$\Lambda_1 = \{c_1, c_2, \ldots, c_{100}\}, c_i \in R$

$\Lambda_2 = \{d_1, d_2, \ldots, d_{100}\}, d_i \in R$

$\boxed{2.2}$ Έστω ένα σύνολο δεδομένων παλινδρόμησης $D = \{X_n, t_n, n=1,\ldots N\}$ τα οποία έχουν παραχθεί ανεξάρτητα βάσει της κανονικής κατανομής, δηλαδή $t_n \sim N(t_n | y(x_n^*, w), b^{-1})$, για κάθε $n=1,2,\ldots N$. Για την εκπαίδευση εφαρμόστε τη τεχνική της μέγιστης πιθανοφάνειας και αποδείξτε την σχέση που πρέπει να ικανοποιεί η βέλτιστη τιμή για την παράμετρο b.

$$\underbrace{N(x|\mu, \sigma^2)}_{\text{γενικός τύπος}} = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$



*output Neural Network

Λύση

$$\sigma^2 = b^{-1} = \frac{1}{b} \xrightarrow{b>0} \sigma = \frac{1}{\sqrt{b}} = \frac{1}{b^{\frac{1}{2}}} = b^{-\frac{1}{2}}$$

Στο πρόχειρο χρήσιμος μετασχηματισμός

Βήμα 1

Υπολογίζουμε την πιθανοφάνεια

$$L = \prod_{n=1}^{N} \mathcal{N}(t_n \mid y(x_n, w), \beta^{-1})$$

Παράδειγμα

$$\frac{\partial (x\cos(x)e^x)}{\partial x} = \frac{\partial (x)}{\partial x}\cos(x)e^x + x\frac{\partial(\cos(x))}{\partial x}e^x +$$

$$+ x\cos(x)\frac{\partial(e^x)}{\partial x} = \ldots\ldots$$

• Βάζω ln ώστε να έχω αθροίσματα αντί πολλαπλ.

$$\frac{\partial}{\partial x}(\ln(x\cos(x)e^x)) = \frac{\partial}{\partial x}(\ln(x) + \ln(\cos(x)) + \ln(e^x)) =$$

$$= \frac{\partial}{\partial x}(\ln(x)) + \frac{\partial}{\partial x}(\ln(\cos(x))) + \frac{\partial}{\partial x}(\ln(e^x))$$

$$\sum_{k=0}^{\infty} x^k \qquad |x|<1 \Rightarrow -1<x<1$$

$$x^k = a_k \to 0$$

Βήμα 2

Υπολογίζουμε το λογάριθμο της πιθανοφάνειας

$$\ln(L) = \sum_{n=1}^{N} \ln(\mathcal{N}(t_n \mid y(x_n, w), \beta^{-1})) =$$

$$= \sum_{n=1}^{N} \ln\left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(t_n - y(x_n, w))^2}\right) =$$

$$\sum_{n=1}^{N} \ln\left(\frac{1}{\frac{1}{\sqrt{b}}\sqrt{2\pi}} e^{-\frac{(t_n - y(x_n,w))^2}{2\cdot\frac{1}{b}}}\right) =$$ εδω ξέρε $\ln(e^x) = x$

$$= \sum_{n=1}^{N} \ln(\sqrt{b}) - \sum_{n=1}^{N} \ln(\sqrt{2\pi}) + \sum_{n=1}^{N} -\frac{b(t_n - y(x_n,w))^2}{2} =$$

$$= \frac{N}{2}\ln(b) - \frac{N}{2}\ln(2\pi) - \frac{b}{2}\sum_{n=1}^{N}(t_n - y(x_n,w))^2$$

Βήμα 3

Βρίσκουμε τη $\frac{\partial}{\partial b}(\ln(L)) = 0 \Rightarrow$

$$\Rightarrow \frac{N}{2}\frac{\partial}{\partial b}(\ln(b)) - \cancel{\frac{\partial}{\partial b}\left(\frac{N}{2}\ln(2\pi)\right)}^{0} - \frac{\partial}{\partial b}(b)\frac{1}{2}\left(\sum_{n=}^{n}(t_n - y(x_n,w))^2\right)$$

$$\Rightarrow \frac{N}{2}\frac{1}{b} + 0 - \frac{1}{2}\sum_{n=1}^{N}(t_n - y(x_n,w))^2 = 0 \Rightarrow$$

$$\Rightarrow \frac{N}{b} = \sum(\ldots) \Rightarrow \frac{1}{b} = \frac{\sum_{n=1}^{N}(t_n - y(x_n,w))^2}{N}$$

τέλος.