ML Q&A Session 6 - Optimization

Thursday, June 18, 2020 10:10 PM

$$L(w) = \frac{1}{N} \cdot \sum_{i=1}^{N} L_i(w)$$

$$\nabla_w L(w) = \frac{1}{N} \boxed{\sum_{i=1}^{N} \nabla L_i(w)} = \frac{1}{N} \cdot \underbrace{\sum_{i=1}^{N} 1}_{=N} = \frac{1}{N} \quad N = 1$$

$$B = \{5, 2, 7, 10, 100\}$$

$$\mathbb{E}\left[\sum_{i \in B} \nabla_w L_i(w)\right] = \nabla_w L(w)$$

$$\nabla_w L_i(w) = 1$$

$$\sum_{i \in B} 1$$

$$\frac{|B|}{\boxed{|B|}\, C} = 1 \qquad \frac{|B|}{C} = N$$

$$\frac{|B|}{|B|} \cdot N = N$$

$$\nabla_w L(w) = \sum_{i=1}^{N} 1 = N$$

$$\nabla L(w) = \boxed{\frac{1}{N}} \sum_{i=1}^{N} \nabla L_i(w) \qquad L_i(w) = (w^T x_i - y_i)^2$$

$$= \frac{1}{N} \sum_{i=1}^{N} 1$$

$$= \frac{1}{N} N = 1$$

$$\tilde{\nabla L}(w) = \sum_{i \in B} \nabla L_i(w) \qquad B = \underbrace{\{1, 5, 10, 20\}}_{M \text{ elements}} \qquad M \ll N$$

$$= C \cdot \sum_{i \in B} 1 = C \cdot M$$

$$\mathbb{E}[\tilde{\nabla L}(w)] = \nabla L(w) \qquad \nabla L_i(w) = 1 \text{ for all } i$$

$$C = \frac{N}{M}$$

$$C \cdot M = N$$

$$C = \frac{N}{M} = \frac{N}{|B|}$$

$$\boxed{C \cdot M = 1 \quad C = \frac{1}{M} = \frac{1}{|B|}}$$

$C = \frac{\ }{M}$ $\quad C \cdot M = N \quad C = \frac{N}{M} = \frac{N}{|S|}$ $\quad \boxed{C \cdot M = 1 \quad C = \frac{1}{M} = \frac{1}{|S|}}$

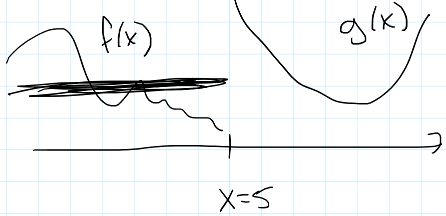
$\nabla_w L(y_i, w^T x_i + b) = \nabla L_i(w)$

if $w^T x_i + b > 0$: predict $y_i = 1$

if $w^T x_i + b < 0$ predict $y_i = -1$

$$\nabla L_i(w) = \begin{cases} 0 & \text{if } y_i(w^T x_i + b) > \varepsilon \\ -\underbrace{y_i}_{\in\{-1,1\}} \underbrace{x_i}_{\in \mathbb{R}^d} & \text{if } y_i(w^T x_i + b) < \varepsilon \end{cases}$$

if $x < 5$ $\quad$ if $x \geq 5$



want $y_i \cdot \mathrm{sign}(w^T x_i + b) > 0$

$$L_i(w) = \max(0, \varepsilon - \overbrace{y_i \cdot (w^T x_i + b)})$$

$$= \begin{cases} 0 & \text{if } y_i(w^T x_i + b) > 0 \\ \varepsilon - y_i(w^T x_i + b) & \text{if } y_i(w^T x_i + b) < 0 \end{cases}$$

$w \in \mathbb{R}^d$

$\nabla L_i(w) \in \mathbb{R}^d$

$x_i \in \mathbb{R}^d \qquad y_i \in \{-1, +1\} \qquad w \in \mathbb{R}^d \qquad b \in \mathbb{R}$

$\delta(x) = \mathrm{sign}(w^T x_i + b)$

$$\underbrace{y_i}\cdot \underbrace{(w^T x_i + b)}_{\in \mathbb{R}} > \varepsilon > 0$$

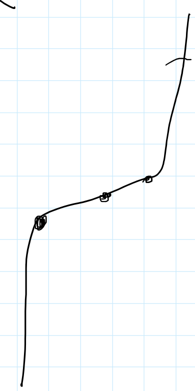
$$L_i(w,b) = \begin{cases} \underbrace{0}_{\in \mathbb{R}} & \text{if } y_i \cdot (w^T x_i + b) > \varepsilon \\ \underbrace{\varepsilon - y_i \cdot (w^T x_i + b)}_{\in \mathbb{R}} & \text{if } y_i (w^T x_i + b) < \varepsilon \end{cases}$$

$\varepsilon - (y_i w^T x_i + \cancel{y_i b})$

$$\nabla_w L_i(w,b) = \begin{cases} \vec{0} \in \mathbb{R}^d & \text{if } y_i \cdot (w^T x_i + b) > \varepsilon \\ -y_i \cdot x_i \in \mathbb{R}^d & \text{else} \end{cases}$$

$f(x, y) \qquad f: \mathbb{R}^2 \to \mathbb{R}$

$f(x,y)$ $\qquad f: \mathbb{R}^2 \to \mathbb{R}$



(x_1, y) $\quad (x_2, y)$

$(-4, 1)$ $\quad (0, 1)$

$$\frac{\partial^2 f(x,y)}{\partial x^2} < 0$$

$\pi \in \mathbb{R}^C$ $\qquad \theta_C \in \mathbb{R}^M$

N samples $\mathcal{D} = \{(x^{(n)}, y^{(n)})\}_{n=1}^{N}$ $\quad x^{(n)} \in \mathbb{R}^{d=2}$ $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

$x = x^{(n)} \quad y = y^{(n)}$ $\qquad y^{(n)} \in \{0,1\}^C$

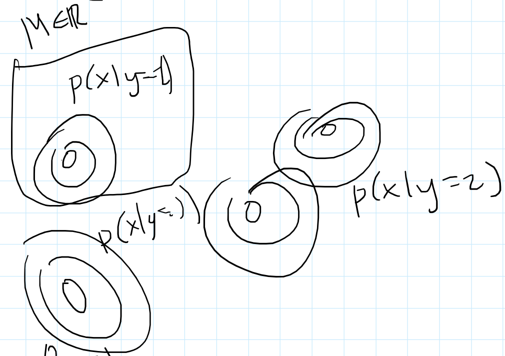
$\Sigma \in \mathbb{R}^{2\times 2}$

$p(x,y) = p(x|y)\,p(y)$ $\qquad [0\ 0\ 1\ 0]$ $\quad C = 4$ $\quad 6$

$\mu \in \mathbb{R}^2$

$[\pi_1\ \pi_2\ \pi_3\ \pi_4]$

$$p(x,y|\theta_1,\dots,\theta_C,\pi) = p(x|y,\theta_1,\dots,\theta_C)\cdot p(y|\pi)$$



$$p(\mathcal{D}\,|\,\theta_{c=1\dots C}, \pi) = \prod_{n=1}^{N} \left(p(x,y\,|\,\theta_1,\dots,\theta_C,\pi) \right)$$

$$\arg\max_{\pi}\ p(\mathcal{D}|\theta_1,\dots,\theta_C,\pi) = \arg\max_{\pi}\ \log p(\mathcal{D}|\theta_1,\dots,\theta_C,\pi)$$

$$\log p(\mathcal{D}|\theta_1,\dots,\theta_C,\pi) = \sum_{n=1}^{N} \log p(x^{(n)}, y^{(n)} \,|\, \dots)$$

$$= \sum_{n=1}^{N} \left(\log p(x^{(n)}|y^{(n)},\theta_1 \dots \theta_C) + \log p(y^{(n)}|\pi) \right)$$

$$= \sum_{n} \log p(y^{(n)}|\pi)$$

$$= \sum_{n} \boxed{\sum_{c=1}^{C} y_c^{(n)} \cdot \log \pi_c}$$

$y^{(n)}$ $\qquad ,0,1,0)$

$$\boxed{p(y^{(n)}|\pi) = \prod^{C} \pi_c^{y_c^{(n)}}}$$

$\sum_n \sum_{c=1}$

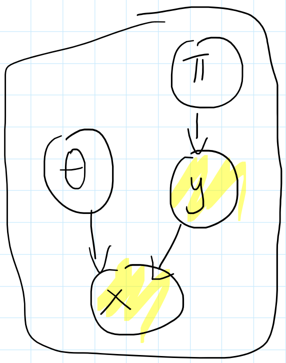
$$\boxed{p(y^{(n)}|\pi) = \prod_{c=1}^{C} \pi_c^{y_c^{(n)}}}$$

$$= \begin{cases} \pi_1 & \text{if } y^{(n)} = 1 = (1,0,0,0) \\ \pi_2 & \text{if } y^{(n)} = 2 = (0,1,0,0) \\ \vdots & (0,0,1,0) \end{cases}$$

$$p(x,y|\theta,\pi) = p(x|y,\theta)\cdot p(y|\pi)$$

$$= p(x|y,\theta,\pi)\cdot p(y|\theta,\pi)$$

$$\boxed{p(a,b|c,d) = p(a|b,c,d)\cdot p(b|c,d)}$$



$y \sim Cat(\pi)$

$x \sim \boxed{p(x|y,\theta)}$ $\boxed{\theta = \{\mu_1, \dots, \mu_C\}}$

$\mathcal{N}(x|\mu_y, I)$

μ_c $y = c$

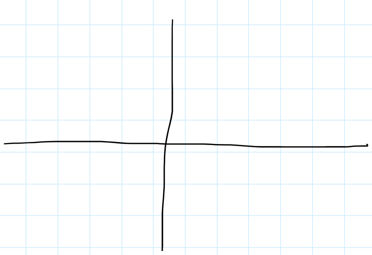
$p(y|x) =$

$p(y=1|x)$ $p(y=0|x)$

$$\delta(x) = \begin{cases} 1 & \text{if } p(y=1|x) \geq 0.5 \\ 0 & \text{if } p(y=1|x) < 0.5 \end{cases}$$

$p(y=1|x) = \sigma(w^T x_i + b)$

$w^T x_i + b > ???$

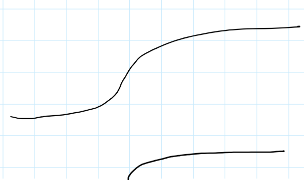


if $\sigma(w^T x_i + b) > 0.5$: $\hat{y}_i = 1$

else: $\hat{y}_i = 0$

for what $w^T x_i + b$ does $\sigma(w^T x_i + b) = 0.5$?

min $e^x + x^2$ $x \to -\infty$



$$\min_{x \in \mathbb{R}} \; e^x + x^2$$

$x \to -\infty$

$e^x \to 0$