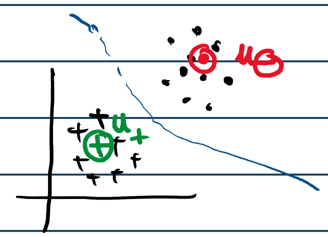
Solutions: Assignment 3

31 March 2024 20:03

Mainak Biswas
PhD student, IISc.



1)

(a)

(i) $$u_0 = \frac{1}{n_0} \sum_{x \in N} x$$

$\Uparrow$ set of -ve data pts

$$u_1 = \frac{1}{n_1} \sum_{x \in P} x$$

(ii) $$D(x) = 0 = \|x - u_0\|^2 - \|x - u_1\|^2$$

$$= \cancel{\|x\|^2} + \|u_0\|^2 - 2u_0^T x - \cancel{\|x\|^2} - \|u_1\|^2 + 2u_1^T x$$

$$= 2(u_1 - u_0)^T x + \left(\|u_0\|^2 - \|u_1\|^2\right)$$

If $D(x) > 0$:

$\hat{y}_{pred} = 1$ $\left[\|x - u_0\|^2 \geq \|x - u_1\|^2\right]$

$\downarrow$ (closer cluster is C_1)

else:

$\hat{y}_{pred} = 0$

(b)

$$\underset{(N \times 1)}{Y_{80 \times 1}} = \underset{(N \times f)}{X_{80 \times 101}} \; \underset{(f \times 1)}{W_{100 \times 1}}$$

$$L(w) = \frac{1}{2}\|Y - Xw\|_2^2 \Rightarrow \frac{\partial L}{\partial w} = X^T(Y - Xw) = 0$$

$L(w) = \frac{1}{2} \| y - Xw \|_2 \Rightarrow \frac{\partial L}{\partial w} = X(y - Xw) = 0$

$$= (X^T X) w = X^T y$$

$$\Downarrow$$

$$\dim(X^T X) = \dim(X) - \dim(\underbrace{N(X^T) \cap C(X)}_{\text{(orthogonal space)}})$$

$$\therefore \operatorname{rank}(X^T X) = \dim(X^T X) - \dim(X)$$

$$\leq \min(80, 101)$$

$\Downarrow$ (101×101) matrix ≤ 80

$\Downarrow$

(not full rank)

$\Downarrow$

(no solution) $\leftarrow$ (not invertible)

prove

2)

(i) $g_1(x) \propto \log P(C_1 | x) \propto \log P(x | C_1) + \log P(C_1)$

$g_2(x) \propto \log P(C_2 | x) \propto \log P(x | C_2) + \log P(C_2)$

$$x | C_1 \sim N\left(\begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}\right)$$

$$x | C_2 \sim N\left(\begin{bmatrix} 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}\right)$$

$\Sigma_1^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1/2 \end{bmatrix}$ $\quad |\Sigma_1| = 2 \quad P(C_1) = 0.6$

$\Sigma_2^{-1} = \begin{bmatrix} 1/2 & 0 \\ 0 & 1/2 \end{bmatrix}$ $\quad |\Sigma_2| = 4 \quad P(C_2) = 0.4$

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$D(x) = g_1(x) - g_2(x) = 0$$

$$= \log \frac{1}{2\pi\sqrt{2}} e^{-\frac{1}{2}\begin{bmatrix} x_1-1 & x_2-2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1/2 \end{bmatrix} \begin{bmatrix} x_1-1 \\ x_2-2 \end{bmatrix}} + \log 0.6$$

$$- \log \frac{1}{4\pi} e^{-\frac{1}{2}\begin{bmatrix} x_1 & x_2+1 \end{bmatrix} \begin{bmatrix} 1/2 & 0 \\ 0 & 1/2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2+1 \end{bmatrix}} - \log 0.4$$

$$= \log \frac{4\pi}{2\pi\sqrt{2}} * \frac{0.6}{0.4} + \frac{1}{4}\left\{ x_1^2 + (x_2+1)^2 - 2(x_1-1)^2 - (x_2-2)^2 \right\}$$

$$= \log \frac{3}{\sqrt{2}} + \frac{1}{4}\left[x_1^2 + (\cancel{x_2^2} + 2x_2 + 1) - (2x_1^2 - 4x_1 + 2) - (\cancel{x_2^2} - 4x_2 + 4) \right]$$

$$= -\log \frac{3}{\sqrt{2}} + \frac{1}{4}\left[x_1^2 - 6x_2 - 4x_1 + 5 \right] = 0$$

$$D(x_1, x_2) = x_1^2 - 4x_1 - 6x_2 + 5 - 4 \log \frac{3}{\sqrt{2}} = 0$$

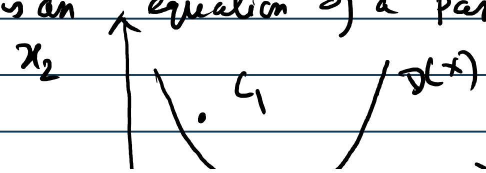
$$6x_2 = (x_1 - 2)^2 + \underbrace{\left\{ 1 - 4\log 3 + 2\log 2 \right\}}_{\approx -2.00}$$

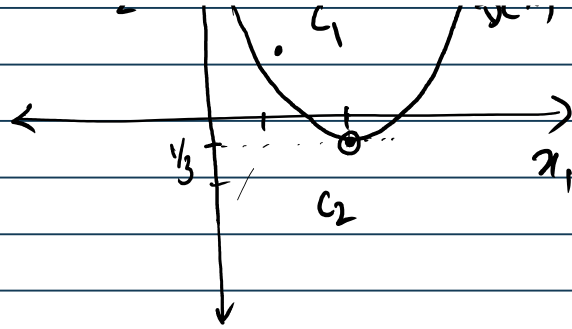
$$(x_1-2)^2 = 6x_2 + 2 \Rightarrow (x_1 - 2)^2 = 4 \cdot \frac{3}{2}\left(x_2 + \frac{1}{3}\right)$$

$$D(x) \equiv (x_1 - 2)^2 = 4\left(\frac{3}{2}\right)\left(x_2 + \frac{1}{3}\right)$$

(ii) Check the implementation in jupyter notebook shared.

But; the above is an equation of a parabola.





3)

<u>given:</u>

$$\alpha_1^*, \alpha_2^* = \underset{\alpha_1, \alpha_2}{\operatorname{argmax}} \quad \omega_0 + \omega_1 \alpha_1 + \omega_2 \alpha_2 - \left[\omega_3 \alpha_1 \alpha_2 + \omega_4 \alpha_1^2 + \omega_5 \alpha_2^2\right]$$

given: $\alpha_1 y_1 + \alpha_2 y_2 = t$.

$$0 \leq \alpha_1, \alpha_2 \leq C$$

(i) From the constraints.

$$\alpha_1 y_1 + \alpha_2 y_2 = t$$

$$\alpha_2 = \frac{1}{y_2}(t - \alpha_1 y_1)$$

$$0 \leq \alpha_1 \leq C \quad , \quad 0 \leq \alpha_2 \leq C$$

substitute value of α_2 here,

$$\Rightarrow \quad 0 \leq \frac{1}{y_2}(t - \alpha_1 y_1) \leq C$$

If $y_1 = y_2$

$\Rightarrow y_1 y_2 = 1$

If $y_1 \neq y_2$

$y_1 y_2 = -1$

$$\Rightarrow y_1 y_2 = 1$$

$$0 \le \frac{y_1}{y_1 y_2}(t - \alpha_1 y_1) \le C$$

$$\Rightarrow 0 \le t y_1 - \alpha_1 \le C$$

$$-t y_1 \le -\alpha_1 \le C - t y_1$$

$$\Rightarrow \overset{0\le}{(t y_1 - C)} \le \alpha_1 \le t y_1$$

$$0 \le \alpha_1 \le C$$

$$\boxed{\max\{0, t y_1 - C\} \le \alpha_1 \le \min(C, t y_1)}$$

(limits)

$$y_1 y_2 = -1$$

$$0 \le \frac{y_1}{-1}(t - \alpha_1 y_1) \le C$$

$$0 \le \alpha_1 - t y_1 \le C$$

$$t y_1 \le \alpha_1 \le t y_1 + C$$

$$0 \le \alpha_1 \le C$$

$$\boxed{\max\{t y_1, 0\} \le \alpha_1 \le \min\{C, t y_1 + C\}}$$

Generalise: $l_1 \le \alpha_1 \le l_2$ (put l_1 and l_2 accordingly)

Objective: $$\mathcal{L}(\alpha_1) = w_0 + w_1 \alpha_1 + w_2\left(\frac{t - \alpha_1 y_1}{y_2}\right) - w_3 \alpha_1 \left(\frac{t - \alpha_1 y_1}{y_2}\right) - w_4 \alpha_1^2 - w_5\left(\frac{t - \alpha_1 y_1}{y_2}\right)^2$$

$$= \left(w_0 + \frac{w_2 t}{y_2} - w_5 t^2\right) + \alpha_1\left(w_1 - w_2 \frac{y_1}{y_2} - \frac{w_3 t}{y_2} + 2 w_5 y_1 t\right) + \alpha_1^2\left(\frac{w_3 y_1}{y_2} - w_4 - w_5\right)$$

$$\mathcal{L}(\alpha_1) = r_2 \alpha_1^2 + r_1 \alpha_1 + r_0 \quad : \text{(quadratic)}$$

$$d(\alpha_1) = \gamma_2 \alpha_1^2 + \gamma_1 \alpha_1 + \gamma_0 \qquad : (\text{quadratic})$$

where; $\quad \gamma_2 = -\left(w_4 + w_5 - w_3 \frac{y_1}{y_2}\right) = \left(w_3 y_1 y_2 - w_4 - w_5\right)$

$$\gamma_1 = \left(w_1 - w_2 y_1 y_2 - w_3 t y_2 + 2 w_5 t y_1\right)$$

$$\gamma_0 = \left(w_0 + w_2 t y_2 - w_5 t^2\right)$$

Analytically :-

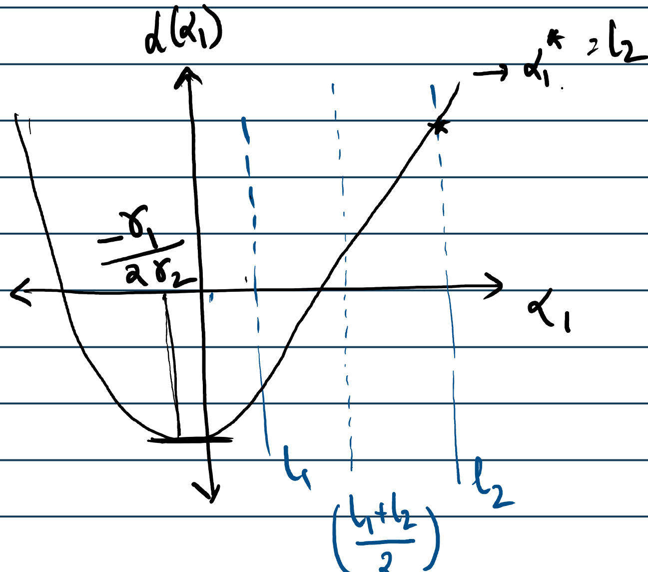
$$\alpha_1^* = \underset{\alpha_1}{\operatorname{argmax}} \; \gamma_2 \alpha_1^2 + \gamma_1 \alpha_1 + \gamma_0$$

$$\text{s.t.} \quad l_1 \leq \alpha_1 \leq l_2$$

Case I :-

$\gamma_2 > 0 \quad$ (upword opening parabola)

(a)



If $\quad -\frac{\gamma_1}{2\gamma_2} \leq \left(\frac{l_1 + l_2}{2}\right)$

$$\alpha_1^* = l_2$$

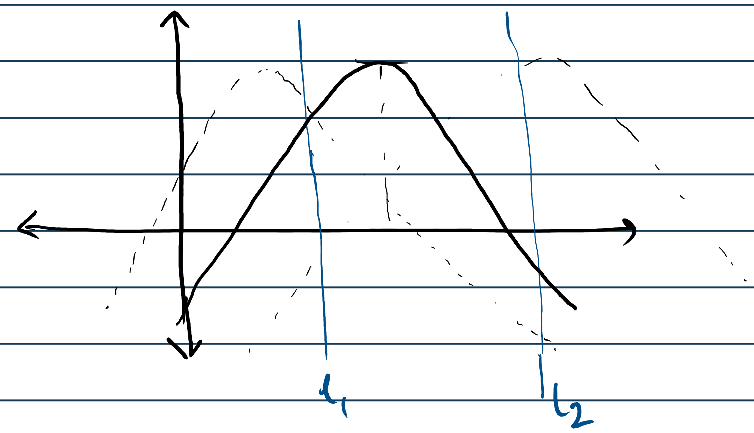
$\alpha_1^* = l_2$

else :

$\alpha_1^* = l_1$

Case II

$\gamma_2 < 0$ (downward opening parabola)



if

$l_1 \leq \frac{-\gamma_1}{2\gamma_2} \leq l_2$

$\alpha_1^* = \frac{-\gamma_1}{2\gamma_2}$

else if $-\frac{\gamma_1}{2\gamma_2} < l_1$

$\alpha_1^* = l_1$

else : $\alpha_1^* = l_2$

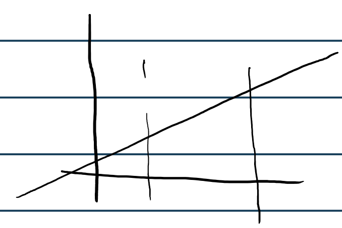
Now you get α_1^*.

$\alpha_2^* = \left(\frac{t - \alpha_1^* y_1}{y_2}\right)$

(Hence solved)

Case III :

if $\gamma_2 = 0$;



line :

If $\gamma_1 > 0$

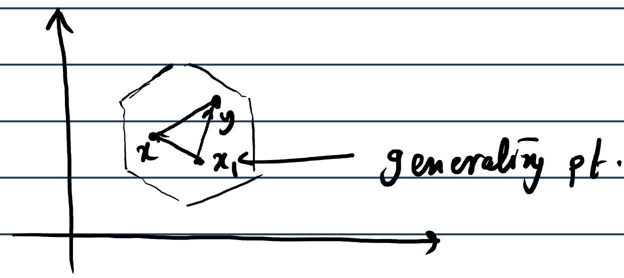
$\alpha_1^* = l_2$

else $\alpha_1^* = l_1$

(Hence solved)

(ii) See the implementation in jupyter notebook.

4)



Let us assign x, y to x_i

i.e $\quad x_i = \underset{x_j \in D}{\operatorname{argmin}} \, d(x, x_j) = \underset{x_j \in D}{\operatorname{argmin}} \, d(y, x_j)$

Each pt has exactly 1 generating function.

Reason: min distance to the data pts is used to assign data points to a neighbour (unique)

To show that the Polygon P_i generated by x_i is Convex:

$$\theta x + (1-\theta) y \in P_i \qquad \forall \, x, y \in P_i$$

in other words:

P.T : $\quad \underset{x_j \in D}{\operatorname{argmin}} \, d\left(\theta x + (1-\theta) y \, ; \, x_j\right) = x_i$

Proof: (Concerning L2 distance)

$$d^2(\theta x + (1-\theta)y, x_j) = \| \theta x + (1-\theta) y - x_j \|_2^2$$

$(\theta + (1-\theta))$

$$= \| \theta (x - x_j) + (1-\theta)(y - x_j) \|^2$$

$$= \theta^2 \| x - x_j \|^2 + (1-\theta) \| y - x_j \|^2$$

$$+ 2\theta(1-\theta)(x - x_j)^T (y - x_j)$$

Consider:

$$(x - x_j)^T (y - x_j)$$

$|x^T y| \geq \|x\| \|y\| \cos\theta$

$$= (x - x_j)^T (y + x - x - x_j)$$

$$= (x - x_j)^T (x - x_j + y - x)$$

$$= \| x - x_j \|^2 + (x - x_j)^T (y - x)$$

$$= \| x - x_j \|^2 + (x + y - y - x_j)^T (y - x)$$

$$= \| x - x_j \|^2 + (y - x_j)^T (y - x) - \| y - x \|^2$$

$$= \| x - x_j \|^2 + (y - x_j)^T (y - x_j + x_j - x) - \| y - x \|^2$$ (independent of x_j)

Cosine rule:

$\Downarrow$

(independent of x_j)

$$\Rightarrow 2(y - x_j)^T (x - x_j) = \| x - x_j \|^2 + \| y - y_j \|^2 - \| y - x \|^2$$

Tnh! Proof: cannot have used

$\Rightarrow 2(y-x_j)^T(x-x_j) = \|x-x_j\|^2 + \|y-y_j\|^2 - \|y-x\|^2$

[Oh! forgot: could have used cosine rule directly]

$\Downarrow$

Derived it instead :(

$\therefore d^2(\theta x + (1-\theta)y, x_j)$

$\Uparrow$ (x_i : maximizes)

$$= \underbrace{\left(\theta^2 + 2\theta(1-\theta)\right)}_{\substack{\theta(2-\theta) \\ >0}} \|x - x_j\|^2 + \underbrace{\left[(1-\theta)^2 + 2\theta(1-\theta)\right]}_{\substack{(1-\theta^2) \\ >0}} \|y - y_j\|^2 + c$$

$\Downarrow$

(x_i maximizes)

Sum of squares maximized when : individually maximized.

$$\underset{x_j}{\operatorname{argmin}}\, d(\theta x + (1-\theta)y, x_j) = x_i$$

$$\therefore \theta x + (1-\theta)y \in P_i$$

[Hence proved : convex set]

5) See the implementation in jupyter.

See : Notes from session 16 to get the derivations for

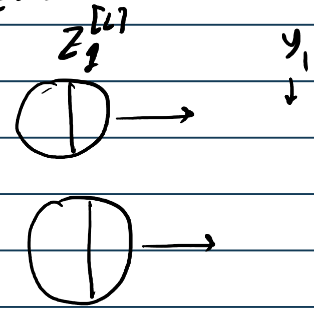
(i) Softmax regression

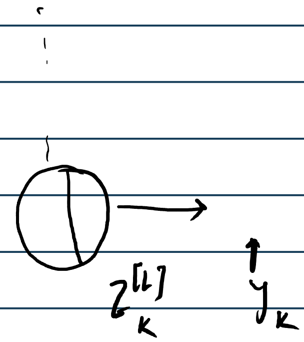
(ii) Logistic regression.

— Rest is just backpropagation — session 21.

Just 1 thing : The $\frac{\partial L}{\partial z^{[L]}}$ is all that needs to be modified.

Last layer :





$$\mathcal{L}(z^{[L]}, y) = \sum_{i=1}^{K} -y_i \log \frac{e^{z_i^{[L]}}}{\sum_{j=1}^{K} e^{z_j^{[L]}}}$$

$$= \sum_{i=1}^{K} -y_i \left\{ z_i^{[L]} - \log \sum_{j=1}^{K} e^{z_j^{[L]}} \right\}$$

$$\frac{\partial L}{\partial z_j} = -\sum_{i=1}^{K} \left\{ y_i \cdot \underbrace{\mathbb{1}_{\{j=i\}}}_{=1 \text{ if } i=j} - \frac{y_i \, e^{z_j^{[L]}}}{\sum_{l=1}^{K} e^{z_l^{[L]}}} \right\}$$

(for 1 pt)

$$= -\left[y_j - \underbrace{\sum_{i=1}^{K} y_i}_{=1} \left(\frac{e^{z_j^{[L]}}}{\sum_{l=1}^{K} e^{z_l^{[L]}}} \right) \right]$$

(indep. of i)

$$= \left[\frac{e^{z_j^{[L]}}}{\sum_{l=1}^{K} e^{z_l^{[L]}}} - y_j \right]$$

vectorised : $\text{softmax}\left(Z^{[L]}\right)_{K \times n_{[L]}} - Y_{K \times n_L}$