Foundations of AI HW #3 B08902016 蔡旭懷

I. Handwritten Part

1. $err(w^T x, y) = (\max(1 - y \cdot w^T x, 0))^2$

$$E_{in}(w) = \frac{1}{N} \sum_{n=1}^{N} err(w^T x_n, y_n)$$

$$\nabla E_{in}(w) = \frac{\partial \left(\frac{1}{N} \sum_{n=1}^{N} err(w^T x_n, y_n) \right)}{\partial w}$$

$$= \frac{1}{N} \sum_{n=1}^{N} \frac{\partial err(w^T x_n, y_n)}{\partial w}$$

$$= \frac{1}{N} \sum_{n=1}^{N} < V_{n,1}, V_{n,2}, V_{n,3}, \dots V_{n,k} >$$

where k = size of vector "w"

$$V_{n,i} = \frac{\partial err(w^T x_n, y_n)}{\partial w_i}$$

$$= \frac{\partial (\max(1 - y_n \cdot w^T x_n, 0))^2}{\partial w_i}$$

① if $1 - y_n \cdot w^T x_n > 0$

$$\Rightarrow V_{n,i} = \frac{\partial ((1 - y_n \cdot w^T x_n)^2)}{\partial w_i}$$

$$= \frac{\partial \left(\left(1 - y_n \cdot \sum_{j=1}^{k} w_j \cdot x_{n,j}\right)^2 \right)}{\partial w_i}$$

$$= \frac{\partial \left(1 - 2y_n \cdot \sum_{j=1}^{k} w_j \cdot x_{n,j} + y_n^2 \cdot \left(\sum_{j=1}^{k} w_j \cdot x_{n,j} \right)^2 \right)}{\partial w_i}$$

$$= -2y_n \cdot x_{n,i} + y_n^2 \cdot \frac{\partial \left(\left(\sum_{j=1}^{k} w_j \cdot x_{n,j} \right)^2 \right)}{\partial w_i}$$

No w_i included

$$= -2y_n \cdot x_{n,i} + \frac{\partial \left(\left(2 \times w_i x_{n,i} \times \sum_{j=1}^{k} w_j x_{n,j} - w_i^2 x_{n,i}^2 + \phi \right) \right)}{\partial w_i} \cdot y_n^2$$

$$= -2y_n \cdot x_{n,i} + 2y_n^2 \times x_{n,i} \times \sum_{j=1, j \neq i}^{k} w_j x_{n,j} + 2y_n^2 x_{n,i}^2 w_i$$

$$= -2y_n \cdot x_{ni} + 2y_n^2 \times x_{n,i} \times \sum_{j=1}^{k} w_j x_{n,j}$$

② if $1 - y_n \cdot w^T x_n \leq 0$

$$\Rightarrow V_{n,i} = \frac{\partial 0}{\partial w_i} = 0$$

$$\therefore \nabla E_{in}(w) = \frac{1}{N}\sum_{n=1}^{N} \langle V_{n,1}, V_{n,2}, \dots V_{n,k} \rangle$$

$$\text{where } \begin{cases} V_{n,i} = -2y_n \cdot x_{n,i} + 2y_n^2 \cdot x_{n,i} \cdot \sum_{j=1}^{k} w_j \cdot x_{n,j}, & \text{if } y_n \cdot w^T x_n < 1 \\ V_{n,i} = 0, & \text{otherwise} \end{cases}$$

#

2. $X_{N\times d} = U_{N\times N} \cdot \Sigma_{N\times d} \cdot (V_{d\times d})^T$

And $U^{-1} = U^T$, $V^{-1} = V^T$

$$\begin{aligned} W_{lin} &= (X^T X)^{-1} X^T y \\ &= (V\Sigma^T U^T \cdot U\Sigma V^T)^{-1} \cdot V\Sigma^T U^T y \\ &= (V\Sigma^T \Sigma V^T)^{-1} \cdot V\Sigma^T U^T y \\ &= V \cdot (\Sigma^T\Sigma)^{-1} \cdot V^T \cdot V\Sigma^T U^T y \\ &= V \cdot (\Sigma^T\Sigma)^{-1} \cdot \Sigma^T U^T y \end{aligned}$$

By Definition, $\Sigma^T\Sigma$ is a $d\times d$ matrix,

And $(\Sigma^T\Sigma)[i,j] = \Sigma[i,j]^2$

And $\Gamma_{d\times N}[i,j] = \begin{cases} \frac{1}{\Sigma[i,j]}, & \text{if } \Sigma[i,j] \neq 0 \\ 0, & \text{otherwise} \end{cases}$

$$\therefore \Sigma^T\Sigma\Gamma = \Sigma^T \Rightarrow (\Sigma^T\Sigma)^{-1} = \Gamma(\Sigma^T)^{-1}$$

As a result, $$\begin{aligned} W_{lin} &= V \cdot (\Sigma^T\Sigma)^{-1} \cdot \Sigma^T U^T y \\ &= V \cdot \Gamma(\Sigma^T)^{-1} \cdot \Sigma^T U^T y \\ &= V\Gamma U^T y \end{aligned}$$

#

3. $N(x|u,I) = \frac{1}{(2\pi)^{D/2}} \times \frac{1}{|I|^{1/2}} \times \exp(-\frac{1}{2}(x-u)^T I^{-1}(x-u))$

$$= \frac{1}{(2\pi)^{D/2}} \times e^{-\frac{1}{2}(x-u)^T \cdot (x-u)}$$

$$u^* = \underset{u \in \mathbb{R}^D}{\arg\max} \prod_{n=1}^{N} p_u(x_n) \quad , \text{ where } \quad p_u(x) = N(x|u,I)$$

$$u^* = \underset{u \in \mathbb{R}^D}{\arg\max} \prod_{n=1}^{N} \frac{1}{(2\pi)^{D/2}} \times e^{-\frac{1}{2}(x_n-u)^T(x_n-u)}$$

$$= \underset{u \in \mathbb{R}^D}{\arg\max} \prod_{n=1}^{N} e^{-\frac{1}{2}(x_n-u)^T(x_n-u)}$$

$$= \underset{u \in \mathbb{R}^D}{\arg\min} \sum_{n=1}^{N} (x_n-u)^T(x_n-u)$$

$$= \underset{u \in \mathbb{R}^D}{\arg\min} \sum_{n=1}^{N} \sum_{k=1}^{D} (x_{n,k}-u_k)^2$$

$$= \underset{u \in \mathbb{R}^D}{\arg\min} \sum_{k=1}^{D} \sum_{n=1}^{N} (x_{n,k}-u_k)^2$$

$$= \underset{u \in \mathbb{R}^D}{\arg\min} \sum_{k=1}^{D} \left(N \times u_k^2 - 2u_k \cdot \sum_{n=1}^{N} x_{n,k} + \sum_{n=1}^{N}((x_{n,k})^2)\right)$$

$$= \underset{u \in \mathbb{R}^D}{\arg\min} \sum_{k=1}^{D} z_k$$

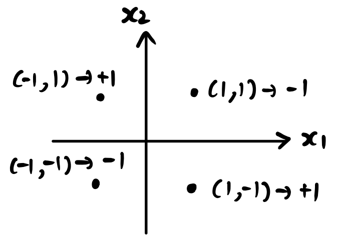
$\forall\ k \in [1, D]$, when $u_k = \frac{-(-2 \cdot \sum_{n=1}^{N} x_{n,k})}{2N} = \frac{\sum_{n=1}^{N} x_{n,k}}{N}$

z_k will have the minimum value

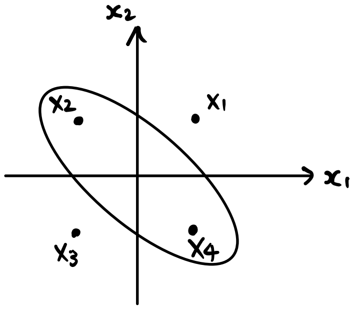
$\therefore\ u^* = \langle u_1, u_2, \dots u_D \rangle$, where $u_i = \underset{n=1}{\overset{N}{\text{Mean}}}(x_{n,i})$

As a result, $u^* = \frac{1}{N} \sum_{n=1}^{N} x_n$ #

4.



Because we can use second-order feature transform ϕ_2, we can use an elliptic curve as the classfier.



The equation of elliptic curve

where $F_1 = X_2$, $F_2 = X_4$, $b = 1$

$$\Rightarrow \sqrt{(x_1-1)^2 + (x_2+1)^2} + \sqrt{(x_1+1)^2 + (x_2-1)^2} = 2\sqrt{3}$$

$$\Rightarrow 2x_1^2 + 2x_2^2 + 2x_1x_2 - 3 = 0$$

let $\tilde{w} = \langle 3, 0, 0, -2, -2, -2 \rangle$

which means $y = \text{sign}(\tilde{w}^T \cdot \phi_2(X))$

$= \text{sign}(-(-3 + 2x_1^2 + 2x_1x_2 + 2x_2^2))$

↘ (just the elliptic curve!)

And $\text{sign}(\tilde{w}^T \cdot \phi_2(X_n)) \equiv y_n$, for $n = 1$ to 4

As a result, $\tilde{w} = \langle 3, 0, 0, -2, -2, -2 \rangle$ is a perceptron such that $y = \text{sign}(\tilde{w}^T \cdot \phi_2(X))$ can separate X_1 to X_4 well.

Besides, $\tilde{w}^T \phi_2(X) = 0$ is an elliptic curve shown below #