

I. Handwritten Part

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

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

$$\begin{aligned} \nabla E_{in}(w) &= \frac{\partial \left(\frac{1}{N} \sum_{n=1}^N \text{err}(w^T x_n, y_n) \right)}{\partial w} \\ &= \frac{1}{N} \sum_{n=1}^N \frac{\partial \text{err}(w^T x_n, y_n)}{\partial w} \\ &= \frac{1}{N} \sum_{n=1}^N \langle v_{n,1}, v_{n,2}, v_{n,3}, \dots, v_{n,k} \rangle \end{aligned}$$

where k = size of vector " w "

$$\begin{aligned} v_{n,i} &= \frac{\partial \text{err}(w^T x_n, y_n)}{\partial w_i} \\ &= \frac{\partial (\max(1 - y_n \cdot w^T x_n, 0))^2}{\partial w_i} \end{aligned}$$

① if $1 - y_n \cdot w^T x_n > 0$

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

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

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

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

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

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

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

$$\textcircled{2} \text{ if } 1 - y_n \cdot w^T x_n \leq 0$$

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

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

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

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

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

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

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

$$\text{And } (\Sigma^T \Sigma)[i, j] = \sum \sigma[i, j]^2$$

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

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

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

$$\begin{aligned}
 3. \quad N(x|u, I) &= \frac{1}{(2\pi)^{D/2}} \times \frac{1}{|I|^{1/2}} \times \exp\left(-\frac{1}{2}(x-u)^T I^{-1} (x-u)\right) \\
 &= \frac{1}{(2\pi)^{D/2}} \times e^{-\frac{1}{2}(x-u)^T (x-u)}
 \end{aligned}$$

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

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

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

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

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

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

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

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

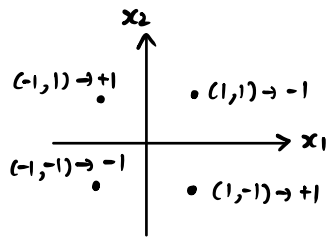
$$\forall k \in [1, D], \text{ when } u_k = \frac{-(-2 \cdot \sum_{n=1}^N x_{n,k})}{2N} = \frac{\sum_{n=1}^N x_{n,k}}{N}$$

z_k will have the minimum value

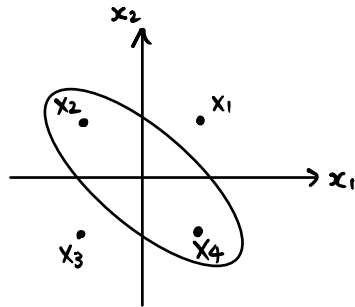
$$\therefore u^* = \langle u_1, u_2, \dots, u_D \rangle, \text{ where } u_i = \text{Mean}(x_{n,i})_{n=1}^N$$

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

4.



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



The equation of elliptic curve

where $F_1 = x_2$, $F_2 = x_4$, $b = 1$

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

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

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

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

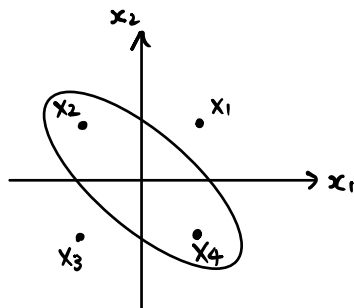
$$= \text{sign}(-(-3 + 2x_1^2 + 2x_1x_2 + 2x_2^2))$$

↳ (just the elliptic curve!)

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

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 #



- Problem 9.

ML Algorithm	Accuracy	F1
Logistic Regression	0.86875	0.73587
Decision Tree	0.80625	0.64520
Random Forest	0.83125	0.62576

- Problem 10.

Number of Trees	F1
5	0.599946
11	0.582149
17	0.642031

The random forest with 17 trees seems to be the best choice.