SET - 2

2.1 Linear Regression

1) Exercise 9.1

→ First, we claim that $|c| = \min_{a \geq 0} a$ s.t. $c \leq a$ & $c \geq -a$ $\forall c \in \mathbb{R}$

Proof:- Case ①: $c \geq 0$

$\Rightarrow |c| = c = \min_{a \geq 0} a$ s.t. $0 \leq c \leq a$ $\Rightarrow c \leq a$ & $c \geq -a$

Case ②: $c < 0$

$\Rightarrow -c > 0$

$\Rightarrow |-c| = -c = \min_{a \geq 0} a$ s.t. $0 < -c \leq a$, i.e. $-a \leq c < 0$

$\therefore$ for any $c \in \mathbb{R}$, $|c| \min_{a \geq 0} a$ s.t. $-a \leq c \leq a$

→ Now consider a vector of auxillary variables $s = (s_1, \ldots, s_m)$. To minimize the emperical risk, we need to minimize the linear objective $\sum_{i=1}^{m} s_i$. And from the above proof we have that $s_i = |w^T x_i - y_i|$

$\Rightarrow -s_i \leq w^T x_i - y_i \leq s_i$

$\Rightarrow w^T x_i - s_i \leq y_i$ and $-w^T x_i - s_i \leq -y_i$ $\forall i \in [0, n]$

→ Let $A \in \mathbb{R}^{2m \times (m+d)}$ be the matrix $A = [X - I_m ; -X - I_m]$, where $X = [x_1, x_2, \ldots, x_m]$

i.e. $A = \begin{bmatrix} x_1^T & \\ \vdots & -I_m \\ x_m^T & \\ -x_1^T & \\ \vdots & -I_m \\ -x_m^T & \end{bmatrix}$

* Define $b \in \mathbb{R}^{2m}$ as

$$b = \begin{bmatrix} y_1 \\ \vdots \\ y_m \\ -y_1 \\ \vdots \\ -y_m \end{bmatrix}$$

and define ~~c as~~ $c \in \mathbb{R}^{m+d}$ as

$$c = \begin{bmatrix} \left.\begin{matrix} 0 \\ \vdots \\ 0 \end{matrix}\right\} d \\ \left.\begin{matrix} 1 \\ \vdots \\ 1 \end{matrix}\right\} m \end{bmatrix}$$

→ Therefore we can represent the ERM problem of linear regression as the following linear program:

$$\min c^T v \quad \text{s.t.} \quad Av \leq b$$

Exercise 9.2

→ 1) Let $X \in \mathbb{R}^{d \times m}$ be defined as $X = [x_1, x_2, \ldots, x_m]$

- The ~~sp~~ rank of X is given by the dimensions of the vector space generated by its columns. i.e.

$$\text{Rank}(X) = \text{dimension of the subspace span}(\{x_1, \ldots, x_m\})$$

- Also, by the Singular Value Decomposition Theorem, the

$$\text{rank}(X) = \text{rank}(XX^T)$$

∴ The set $X = \{x_1, \ldots, x_m\}$ spans $\mathbb{R}^d$ iff $\text{rank}(XX^T) = d$

- If $\text{rank}(XX^T) = d$, then the matrix XX^T is non-singular and therefore invertible.

∴ $x_1, x_2, \ldots, x_m$ spans $\mathbb{R}^d$ iff $A = XX^T$ is invertible

2) The ERM problem is represented as

$$\min_{W \in \mathbb{R}^{n \times d}} \frac{1}{m} \sum_{i=1}^{m} \| W^T \phi(x_i) - y_i \|^2$$

$$= \min_{W \in \mathbb{R}^{n \times d}} \frac{1}{m} \sum_{i=1}^{m} \sum_{j=1}^{d} (W_j^T \phi(x_i) - y_{ij})^2 \text{, where } W = [W_j] \text{ \& } y_{ij} = [y_i]$$

$$= \min_{W \in \mathbb{R}^{n \times d}} \frac{1}{m} \sum_{j=1}^{d} \sum_{i=1}^{m} (W_j^T \phi(x_i) - y_{ij})^2 \quad \text{as } i \text{ and } j \text{ are independent in the expression}$$

$$= \min_{W \in \mathbb{R}^n} \frac{1}{m} \sum_{i=1}^{m} (W_1^T \phi(x_i) - y_{i1})^2 + \min_{W \in \mathbb{R}^n} \frac{1}{m} \sum_{i=1}^{m} (W_2^T (\phi(x_i)) - y_{i2})^2 + \ldots$$

$$\ldots + \min_{W \in \mathbb{R}^n} \frac{1}{m} \sum_{i=1}^{m} (W_d^T \phi(x_i) - y_{id})^2 \quad \text{(As they are linearly independent)}$$

→ ~~We can split~~ Therefore the ERM problem on W can be represented as d individual linear regression problems.

Hence Proved

2.1.3)

- For $\Phi(x) = x$, the explained variance is approximately 0.4063
- For $\Phi(x) = [1\ x\ x^2]^T$, the explained variance is approximately 0.4032

2.2)

a) Using scikit Perceptron class

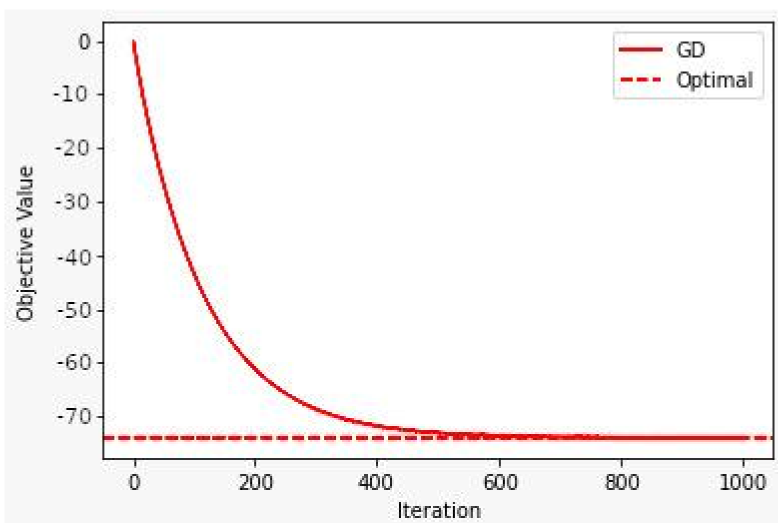
- For $\Phi(x) = x$, the classification accuracy is approximately 81
- For $\Phi(x) = [1\ x\ x^2]^T$, the classification accuracy is approximately 80

b) Logistic Regression classification

- For $\Phi(x) = x$, the classification accuracy is approximately 82
- For $\Phi(x) = [1\ x\ x^2]^T$, the classification accuracy is approximately 84

2.3)

b)



c)