Machine-Learning 2 · EE660

HW

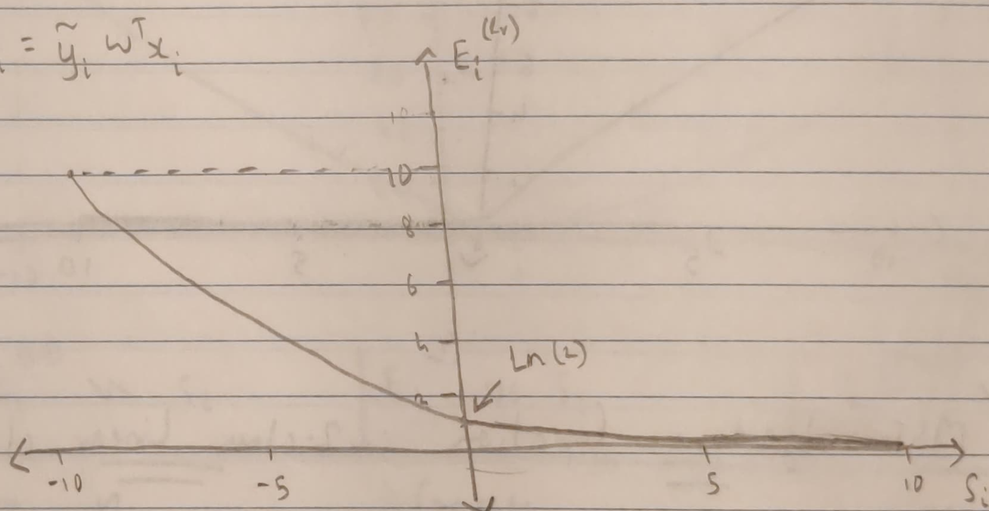
HW 1

↳ Hardik Prajapati (2678294168)

1. a) Logistic Regression based on MLE:-

$$E_i^{(Lr)} = \ln\left[1 + \exp\{-\tilde{y}_i w^T x_i\}\right]$$

$S_i = \tilde{y}_i w^T x_i$

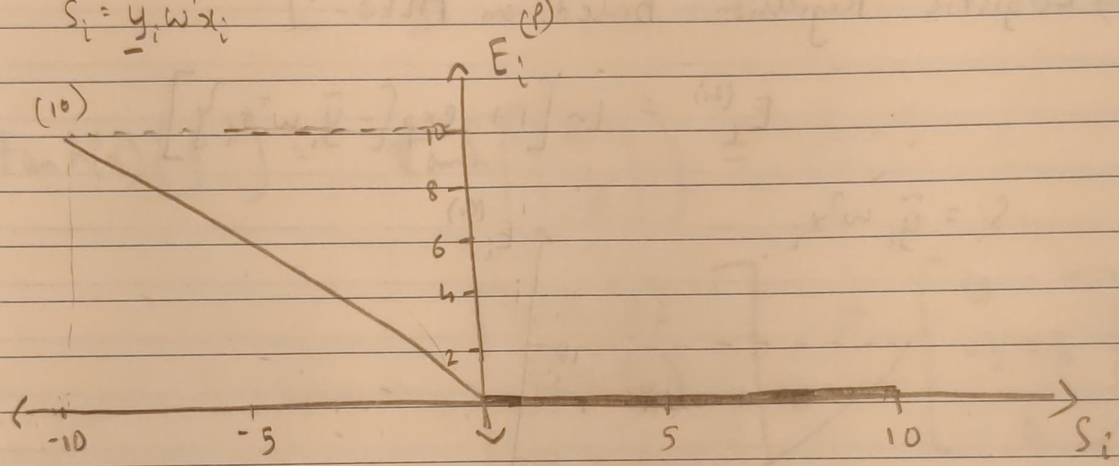


b. 2-class Linear Perceptron Learning

$$J(w) = -\sum_{i=1}^{N} [[\tilde{\underline{y}}_i w^T x_i \leq 0]] \tilde{\underline{y}}_i w^T x_i = \sum_{i=1}^{N} E_i^{(p)}$$

$$E_i^{(p)} = \begin{cases} -\tilde{\underline{y}}_i w^T x_i & \text{if } \tilde{\underline{y}}_i w^T x_i \leq 0 \\ 0 & \text{if } \tilde{\underline{y}}_i w^T x_i > 0 \end{cases}$$

$s_i = \tilde{\underline{y}}_i w^T x_i$



c. MSE objective function :- 2-class Linear classification

$$MSE = \frac{1}{N} \sum_{i=1}^{N} [w^T x_i - b_i]^2 = \sum_{i=1}^{N} E_i^{(MSE)}$$

$$E_i^{(MSE)} = [w^T x_i - b_i]^2$$

$$= [w^T x_i - \tilde{\underline{y}}_i]^2$$

$$= (w^T x_i)^2 + (\tilde{\underline{y}}_i)^2 - 2\tilde{\underline{y}}_i w^T x_i$$

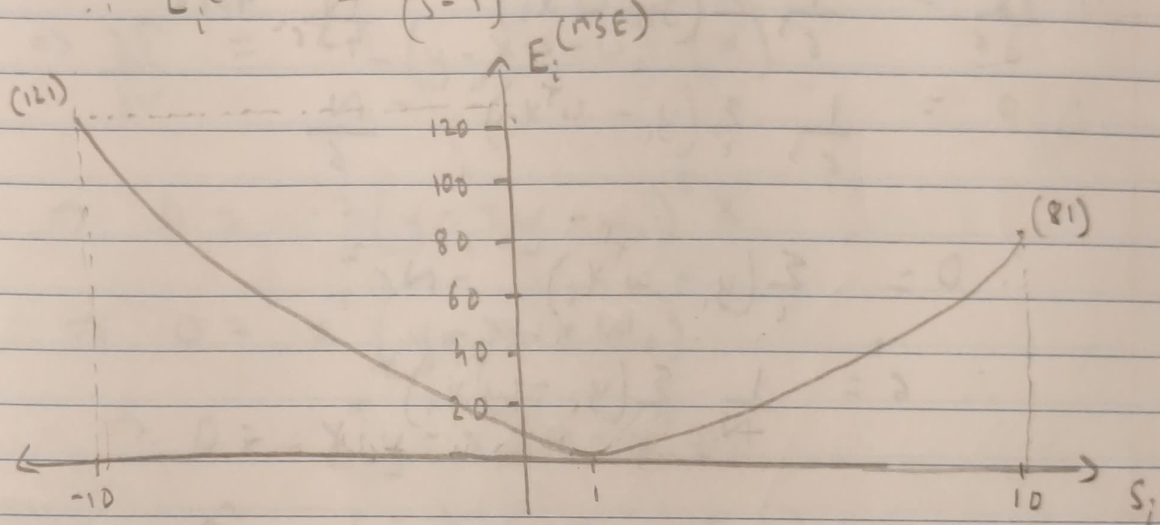
$$\therefore E_i^{(MSE)} = (w^T x_i)^2 + 1 - 2\,\underline{\tilde{y}}_i w^T x_i$$

$$\therefore E_i^{(MSE)} = (\underline{\tilde{y}}_i w^T x_i)^2 + 1 - 2\,\underline{\tilde{y}}_i w^T x_i$$

Let $s = \underline{\tilde{y}}_i w^T x_i$

$$\therefore E_i^{(MSE)} = s^2 + 1 - 2s$$

$$\therefore E_i^{(MSE)} = (s-1)^2$$



d.

$E_i^{(LR)}$ vs S_i	$E_i^{(P)}$ vs S_i	$E_i^{(MSE)}$ vs S_i
1) Convex 2) Error function merges to zero as we move away from boundary (for correctly classified samples) 3) For wrongly classified samples, points near boundary (on wrong side) has less error cost	1) Convex 2) Error function merges to zero as we move away from boundary. (For correctly classified samples) 3) For wrongly classified samples, points near boundary (on wrong side) has less error cost.	1) Convex 2) Error function increases as we move away from boundary (Even for rightly classified samples) 3) For wrongly classified samples, points near boundary (on wron side) has high error cost compared to previous two

3. MLE for σ^2 for Linear Regression

To show:- $\hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^{N} (y_i - x_i^T \hat{\omega})^2$

$$l = \underset{\omega, \sigma^2}{NLL} = -\frac{1}{2\sigma^2} \sum (y_i - \omega^T x_i)^2 - \frac{N}{2} \log(2\pi\sigma^2)$$

$$\therefore \frac{\partial l}{\partial \sigma} = \frac{1}{\sigma^3} \sum_i (y_i - \omega^T x_i)^2 - \frac{N}{4\pi\sigma^2} 4\pi\sigma$$

$$\therefore 0 = \frac{1}{\sigma^3} \sum_i (y_i - \omega^T x_i)^2 - \frac{N}{\sigma}$$

$$\therefore 0 = \sum_i (y_i - \omega^T x_i)^2 - N\sigma^2$$

$$\therefore \sigma^2 = \frac{1}{N} \sum_i (y_i - \omega^T x_i)^2$$

$\frac{\partial l}{\partial \omega} = 0$ will give $\omega = \hat{\omega}$.

Substituting $\omega = \hat{\omega}$ in above equation yields,

$$\therefore \sigma^2 = \frac{1}{N} \sum_i (y_i - \hat{\omega}^T x_i)^2$$

$$\therefore \sigma^2 = \frac{1}{N} \sum_i (y_i - x_i^T \hat{\omega})^2$$ $\rightarrow$ {Since $\omega^T x_i$ is scalar

Hence, proved.

4. MLE for w_0 (offset term) in Linear Regression

$$J(w, w_0) = \frac{1}{N} \sum_{i=1}^{N} \left(y_i - (w_0 + w^T x_i)\right)^2$$

$$\Rightarrow J(w, w_0) = \frac{1}{N} \sum_{i=1}^{N} \left(y_i^2 - 2y_i(w_0 + w^T x_i) + (w_0 + w^T x_i)^2\right)$$

$$= \frac{1}{N} \sum_{i=1}^{N} \left(y_i^2 - 2y_i w_0 - \underbrace{2y_i w^T x_i}_{-2y_i x_i^T w} + w_0^2 + \underbrace{2w_0 w^T x_i}_{2w^T x_i w_0} + w^T x_i x_i^T w\right)$$

$$\Rightarrow \frac{\partial J}{\partial w_0} = \frac{1}{N} \sum_{i=1}^{N} \left(-2y_i + 2w_0 + 2w^T x_i\right)$$

$$\Rightarrow 0 = -\frac{1}{N} \sum_{i=1}^{N} y_i + \frac{1}{N} \sum_{i=1}^{N} w_0 + \frac{1}{N} \sum_{i=1}^{N} w^T x_i$$

$$\Rightarrow 0 = -\frac{1}{N}\sum_{i=1}^{N} y_i + \frac{w_0}{N}(N) + \frac{1}{N}\sum_{i=1}^{N} w^T x_i$$

$$\Rightarrow w_0 = \frac{1}{N}\sum_{i=1}^{N}\{y_i\} - \frac{1}{N}\sum_{i=1}^{N} w^T x_i$$

OR.

$$\Rightarrow w_0 = \frac{1}{N}\sum_{i=1}^{N} y_i - \frac{1}{N}\sum_{i=1}^{N} x_i^T w = \bar{y} - \bar{x}^T w$$

Now,

$$J(w, w_0) = (y - Xw - w_0 1_n)^T (y - Xw - w_0 1_n)$$

Substituting, $\{w_0 = \bar{y} - \bar{x}^T w\}$ in above,

$$\Rightarrow J(w, w_0) = (y - Xw - \bar{y} + \bar{x}^T w)^T (y - Xw - \bar{y} + \bar{x}^T w)$$

$$\Rightarrow J(w, w_0) = ((y - \bar{y}) - (Xw - \bar{x}^T w))^T ((y - \bar{y}) - (Xw - \bar{x}^T w))$$

$$\Rightarrow J(w, w_0) = (y_c - X_c w)^T (y_c - X_c w)$$

Taking Derivative,

$$\Rightarrow \frac{\partial J}{\partial w} = 2(y_c - X_c w)^T (-X_c)$$

$$\Rightarrow 0 = (y_c - X_c w)^T X_c$$

$$\Rightarrow 0 = y_c^T x_c - w^T x_c^T x_c$$

$$\Rightarrow w^T x_c^T x_c = y_c^T x_c$$

$$\Rightarrow x_c^T x_c w = x_c^T y_c$$

$$\Rightarrow w = (x_c^T x_c)^{-1} x_c^T y_c$$ → (Considering inverse exists)

//