Name - TEJAS ACHARYA

EE-559

HOMEWORK - 04

16-06-2023

① 2-Class Perceptron Learning, Criterion Function

$$J(\underline{w}) = -\sum_{n=1}^{N} [\![\underline{w}^T z_n x_n \le 0]\!] \, \underline{w}^T z_n x_n$$

(a) Using max{ }

$$J(\underline{w}) = \sum_{n=1}^{N} \max\{0, -\underline{w}^T z_n x_n\}$$

(b) • 0 is convex

• $-\underline{w}^T z_n x_n$ is convex

Using Pointwise Max property of convex Function

$\max\{0, -\underline{w}^T z_n x_n\}$ is convex

$$J(\underline{w}) = \sum_n \underbrace{\max\{0, -\underline{w}^T z_n x_n\}}_{\text{convex}}$$

$$= \max\{0, -\underline{w}^T z_0 x_0\} + \max\{0, -\underline{w}^T z_1 x_1\} + \dots$$

From Non-negative weighted summation.

$\boxed{J(\underline{w}) \text{ is also convex}}$

(2) Assumptions

- Class, $C = 2$
- Perceptron with Margin ; Let the Margin be $\underline{b}$; $b > 0$
- Basic Sequential G.D
- Fixed Increment $\Rightarrow \eta(i) = \eta > 0$

Use reflected points, $z_n \underline{x}_n$, $n = 1, 2, \dots N$

Consider $\eta = 1$

$$\Rightarrow \quad z_n \underline{x}_n' = z_n \eta \underline{x}_n = z_n \underline{x}_n$$

Algo $\to$ Randomize Dataset

$\underline{w}(0)$ = Random initialization

For $m = 1, 2, \dots$ epochs

For $n = 1, 2, \dots N$

$$\underline{w}(i+1) = \underline{w}(i) + z_i \underline{x}_i \, [\![\underline{w}^T(i) z_n \underline{x}_n \le b]\!]$$

Consider $z^i \underline{x}^i$ points misclassified in i^{th} iteration

$$\Rightarrow [\![\underline{w}^T(i) z^i \underline{x}^i \le b]\!] = 1$$

Algo.

- Randomize Data
- $\underline{w}(0)$ = Random initialization

For $m = 1, 2, \dots$ epochs

For $n = 1, 2, \dots N$

$$\underline{w}(i+1) = \underline{w}(i) + z^i \underline{x}^i_n$$

Let $\hat{\underline{w}}$ be a solution

$$\Rightarrow \hat{\underline{w}}^T z_n \underline{x}_n > b, \ \forall n$$

Consider $a > 0 \Rightarrow a\hat{\underline{w}}$ is also a sol^n

$$\Rightarrow a \hat{\underline{w}}^T z_n \underline{x}_n > b$$

(2) Error measure on $\underline{w}(i)$ w.r.t soln $a\hat{\underline{w}}$

$$E_{\underline{w}}(i) = \| \underline{w}(i) - \underset{\substack{\uparrow \\ \text{Adjustable}}}{a}\, \hat{\underline{w}} \|_2^2$$

<u>To Prove convergence</u> we must show $E_{\underline{w}}(i)$ decreases with iteration

To show $\boxed{E_{\underline{w}}(i+1) \le E_{\underline{w}}(i)}$

We know $\underline{w}(i+1) = \underline{w}(i) + z^i \underline{x}^i$

$\left.\begin{array}{l} -a\hat{\underline{w}} \text{ on both} \\ \text{sides} \end{array}\right\} \Rightarrow \quad \underline{w}(i+1) - a\hat{\underline{w}} = \left(\underline{w}(i) - a\hat{\underline{w}}\right) + z^i \underline{x}^i \quad , a > 0$

$(\text{L-2 norm})^2$ on both sides $\Rightarrow$

$$\| \underline{w}(i+1) - a\hat{\underline{w}} \|_2^2 = \| (\underline{w}(i) - a\hat{\underline{w}}) + z^i \underline{x}^i \|_2^2$$

$$= \| \underline{w}(i) - a\hat{\underline{w}} \|_2^2 + 2\left[\underline{w}(i) - a\hat{\underline{w}}\right]^T z^i \underline{x}^i + \| z^i \underline{x}^i \|_2^2$$

$$E_{\underline{w}}(i+1) = E_{\underline{w}}(i) + \overset{\text{can be dropped}}{\underset{\le b}{2\,\underline{w}^T(i)\, z^i \underline{x}^i}} - 2a \underbrace{\hat{\underline{w}}^T z^i \underline{x}^i}_{> b} + \underbrace{\| z^i \underline{x}^i \|_2^2}_{> 0}$$

Let $t^2 \triangleq \max_j \| x_j \|_2^2$

$c \triangleq \min_j \{ \hat{\underline{w}}^T z_j x_j \} > b$

$$E_{\underline{w}}(i+1) \le E_{\underline{w}}(i) - 2ac + t^2$$

Choose $a = \dfrac{t^2}{c} > 0$

$$E_{\underline{w}}(i+1) \le E_{\underline{w}}(i) - t^2$$

So each iteration reduces E_w by at least $\underline{\underline{t^2}}$

③ Forcing Argument

$$0 \le E_{\underline{w}}(i+1) \le E_{\underline{w}}(i) - t^2 \quad \forall i$$

For some i_0 we would have $E_{\underline{w}}(i_0) < t^2$

$$\Rightarrow \quad 0 \le E_{\underline{w}}(i_0+1) \le E_{\underline{w}}(i_0) - t^2 < 0$$

- Not Possible
- Stop iterations at $i_0 - 1$ or earlier.
- Algo converges at $\underline{i_0 - 1}$ iteration or earlier.

---

④ Linear Regression, $\hat{y}_n = \underline{w}^T \underline{x}_n$ [Augmented space]

Mean Squared Criterion Function $\Big\}$ $J(\underline{w}) = \frac{1}{N} \sum_{n=1}^{N} \left[\hat{y}_n - y_n\right]^2 = \frac{1}{N} \sum_{n=1}^{N} \left[\underline{w}^T \underline{x}_n - y_n\right]^2$

ⓐ $\underline{w} = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_D \end{bmatrix}$, $\underline{\underline{X}} = \begin{bmatrix} \underline{x}_1^T \\ \underline{x}_2^T \\ \vdots \\ \underline{x}_N^T \end{bmatrix}$; $\underline{x} = \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_D \end{bmatrix}$; $\underline{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$

$$\underline{\underline{X}}\,\underline{w} = \begin{bmatrix} \underline{x}_1^T \underline{w} \\ \underline{x}_2^T \underline{w} \\ \vdots \\ \underline{x}_N^T \underline{w} \end{bmatrix} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_N \end{bmatrix} = \hat{\underline{y}}$$

$J(\underline{w})$ can be rewritten as $\rightarrow$ $J(\underline{w}) = \frac{1}{N} \left\| \underline{\underline{X}}\,\underline{w} - \underline{y} \right\|_2^2$

$$= (\underline{\underline{X}}\,\underline{w} - \underline{y})^T (\underline{\underline{X}}\,\underline{w} - \underline{y})$$

4 (a)

$$\nabla_{w} J(w) = \frac{1}{N}\left[X^T (X w - y) + (X w - y)^T X \right]$$

$$= \left[X^T X w - X^T y + (X w)^T X - y^T X \right]$$

$$= \left[X^T X w - X^T y + w X^T X - y^T X \right]$$

2nd order Derivative

$$\nabla_{w} \left[\nabla_{w} J(w) \right] = \left[X^T X + X^T X \right]$$

$$H_{w}(J(w)) = 2 X^T X$$

Show H_w to be positive semi-definite

Take $a \neq 0$

Show $a^T H a \geq 0$

$$a^T (2 X^T X) a = 2 a^T X^T X a = 2 (X a)^T (X a)$$

$$= 2 \| X a \|_2^2 \geq 0 \quad \forall a$$

$\therefore$ $a^T H a \geq 0$

$\therefore$ $H_w(J(w))$ is positive - semi definite.

$\therefore$ $J(w)$ is a Convex Function

4 (b) (i) Lasso Regression Criterion Function

$$J(\underline{w}) = \frac{1}{N}\sum_{n=1}^{N}[\hat{y}_n - y_n]^2 + \lambda\sum_{j=1}^{D+1}|w_j| \quad , \lambda \geq 0$$

In Matrix Form,

$$J(\underline{w}) = \underbrace{(\underline{\underline{X}}\,\underline{w} - \underline{y})^T(\underline{\underline{X}}\,\underline{w} - \underline{y})}_{\text{From } \boxed{4(a)}} + \lambda\|\underline{w}\|_1$$

(ii) Is $\lambda\|\underline{w}\|_1$ convex ?

a(i) $\|\underline{w}\|_1 \geq 0$ for all $\underline{w}$, b(i) $\|\lambda\underline{w}\|_1 = |\lambda|\,\|\underline{w}\|_1 = \lambda\|\underline{w}\|_1$ as $\lambda \geq 0$

Consider $\underline{a}$, $\underline{b}$ & $0 \leq c \leq 1$

$\Rightarrow$ c(i) $\|\underline{a} + \underline{b}\|_1 \leq \|\underline{a}\|_1 + \|\underline{b}\|_1$ $\rightarrow$ Triangle Inequality

Combining a(i), b(i), c(i),

$$\|c\underline{a} + (1-c)\underline{b}\|_1 \leq \|c\underline{a}\|_1 + \|(1-c)\underline{b}\|_1$$

$$\|c\underline{a} + (1-c)\underline{b}\|_1 \leq |c|\,\|a\|_1 + |(1-c)|\,\|b\|_1$$

$0 \leq c \leq 1 \Rightarrow$ $\boxed{\|c\underline{a} + (1-c)\underline{b}\|_1 \leq c\|a\|_1 + (1-c)\|b\|_1}$

Satisfies convex function condition.

i.e. if $\boxed{f(c\underline{a} + (1-c)\underline{b}) \leq cf(\underline{a}) + (1-c)f(\underline{b})}$, then $f()$ is convex

$\therefore$ $\|\underline{w}\|_1$ is convex

4 (ii) $\lambda \|\underline{w}\|_1$ is convex from Non-Negative Weighted Summation as $\lambda \geq 0$

(iii) $J(\underline{w}) = \underbrace{(\underline{\underline{X}}\,\underline{w} - \underline{y})^T (\underline{\underline{X}}\,\underline{w} - \underline{y})}_{\text{Convex from 4(a)}} + \underbrace{\lambda \|\underline{w}\|_1}_{\text{Convex from 4(b)(ii)}}$

$\therefore$ Using Non-Negative Weighted Summation

$J(\underline{w})$ is also convex