Our hypothesis consists of an hyperplane whose equation is

$$w^T x + b = 0$$

For some iteration t, let w_t, b_t be the w and b.
For that iteration, let Δw_t, Δb_t be the gradients

$$\therefore \left.\begin{aligned} w_{t+1} &= w_t - \alpha \Delta w_t \\ b_{t+1} &= b_t - \alpha \Delta b_t \end{aligned}\right\} \alpha \text{ is the step size}$$

Therefore, Our hyperplane becomes:

$$w_{t+1}^T x + b_{t+1} = 0$$

$$(w_t - \alpha \Delta w_t)^T x + (b_t - \alpha \Delta b_t) = 0$$

$$w_t^T x - \alpha \Delta w_t^T x + b_t - \alpha \Delta b_t = 0$$

$$\underbrace{w_t^T x + b_t}_{=0} - \alpha \Delta w_t^T x - \alpha \Delta b_t = 0$$

$\hookrightarrow$ this is our previous ~~of~~ hyperplane

$$\therefore \quad -\alpha \Delta w_t^T x - \alpha \Delta b_t = 0$$

$$\therefore \quad -\alpha (\Delta w_t^T x + \Delta b_t) = 0$$

$$\Leftrightarrow \Delta w_t^T x + \Delta b_t = 0 \qquad (\because \alpha \neq 0)$$

$\therefore$ No matter what α (step size) we select, we always iteratively predict the same hyperplane.

$\therefore$ The rat of convergence do not change with α.

Problem 2

1. Is the data linearly seperable

(a) $\phi(x_1, x_2) = \begin{bmatrix} x_1 + x_2 \\ x_1 - 2x_2 \end{bmatrix}$

x_1	x_2	x_1' $x_1 + x_2$	x_2' $x_1 - 2x_2$	label
-1	-1	-2	1	$+$
1	1	2	-1	$+$
-1	1	0	-3	$-$
1	-1	0	3	$-$

$\therefore$ We need to find a line
$$w_1 x_1' + w_2 x_2' + b = 0$$
s.t.

$$b + -2w_1 + w_2 > 0 \qquad\qquad b - 3w_2 < 0$$
$$b + 2w_1 - w_2 > 0 \qquad\qquad b + 3w_2 < 0$$
$$(+)\ \overline{\qquad\qquad\qquad} \qquad\qquad (+)\ \overline{\qquad\qquad\qquad}$$
$$2b > 0 \quad -(1) \qquad\qquad 2b < 0 \quad -(2)$$

From (1) and (2);
$$\boxed{b = 0.}$$

now, $-2w_1 + w_2 > 0 \Rightarrow w_2 > 2w_1$

$2w_1 - w_2 > 0 \Rightarrow 2w_1 > w_2$

$\omega_2 > 2\omega_1$ & $\omega_1 > 2\omega_2$

cannot be satisfied with any values of ω_1, ω_2

$\therefore$ Linear seperator does not exist.

(b) $\phi(x_1, x_2) = \begin{bmatrix} x_1^2 \\ x_2^2 \\ x_1 x_2 \end{bmatrix}$

This set contains a linear seperator.

I have solved it using code and the seperator is given by

$\Rightarrow \omega_1 x_1^2 + \omega_2 x_2^2 + \omega_3 x_1 x_2 + b = 0$

where $\begin{bmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \end{bmatrix} = \begin{bmatrix} \quad \end{bmatrix}$

and $b =$

(c) $\phi(x_1, x_2) = \begin{bmatrix} \exp(x_1) \\ \exp(x_2) \end{bmatrix}$

$\therefore$ We want $w_1 e^{x_1} + w_2 e^{x_2} + b = 0$

for

$w_1 e^{-1} + w_2 e^{-1} + b > 0$ — ①

$w_1 e + w_2 e + b > 0$ — ②

$w_1 e^{-1} + w_2 e + b < 0$ — ③

$w_1 e + w_2 e^{-1} + b < 0$ — ④

Add ① and ②;

$w_1 (e^{-1} + e) + w_2 (e + e^{-1}) + 2b > 0$ — ⑤

Add ③ and ④,

$w_1 (e^{-1} + e) + w_2 (e^{-1} + e) + 2b < 0$ — ⑥

From ⑤ and ⑥,

We conclude that a linear seperator is not possible.

Problem 2: Q2

Polynomial regression for 2-D data points.

The data in 2-d is of the form (x, y).

Thus, for polynomial regression, we need a K-degree polynomial of x for which the output is y.

$\therefore$ Our hypothesis function:

$$w_K x^K + w_{K-1} x^{K-1} + \ldots\ldots + b x^0 = \underbrace{y'}_{\text{Predictions.}}$$

In vector notation;

$$\underline{f(x) = \omega^T X + b}$$

where $X = \underbrace{\begin{bmatrix} x^K \\ x^{K-1} \\ \vdots \\ \vdots \\ x \end{bmatrix}}_{\text{Feature Vectors}}$ $\quad \omega = \begin{bmatrix} \omega^K \\ \omega^{K-1} \\ \vdots \\ \omega_1 \end{bmatrix}$

Feature vector X is generated from our input x.

The loss function: mean square error

$$L(f) = \frac{1}{M} \sum_{m} \left(f(x^m) - y^m \right)^2$$

$$L(f) = \frac{1}{M} \sum_{m} \left((w^T x^m + b) - y^m \right)^2$$

And the objective function is given by

$$\min_{w,b} L(f) = \min_{w,b} \frac{1}{M} \sum_{m} \left((w^T x^m + b) - y^m \right)^2$$

Given the objective, we will use gradient descent to compute the w, b for minimum loss

$\therefore$ Gradient descent

$$\cancel{\quad} \frac{\partial L}{\partial w} = \frac{2}{M} \sum_{m} \left((w^T x^m + b) - y^m \right) \cdot x^m$$

$$\frac{\partial L}{\partial b} = \frac{2}{M} \sum_{m} \left((w^T x^m + b) - y^m \right)$$

Iteratively compute:

$$w_{t+1} := w_t - \alpha \frac{\partial L}{\partial w}$$

$$b_{t+1} := b_t - \alpha \frac{\partial b}{\partial b}$$

Untill, $\frac{\partial L}{\partial w} \simeq 0$ and $\frac{\partial L}{\partial b} \simeq 0$.

Time Complexity (K degree polynomial, m data points)

For a K degree polynomial, the predictions will take $O(Km)$ time. And the gradient computing using the predictions also takes $O(Km)$ time.

$\therefore$ Time Complexity $= O(Km) + O(Km) = \underline{O(Km)}$

Problem 3 Exponential Regression.

) **Input** $(x, y) \in \mathbb{R}^2$

M input data points.

Hypothesis

$$f(x) = \exp(ax+b)$$

loss

$$L(f) = \frac{1}{M} \sum_m \left(\exp(ax^m+b) - y^m\right)^2$$

objective

$$\min_{a,b} L(f) = \min_{a,b} \frac{1}{M} \sum_m \left(\exp(ax^m+b) - y^m\right)^2$$

We minimize this function using gradient descent.

Gradient Descent

$$\min_{a,b} L(f) = \min_{a,b} \frac{1}{M} \sum_{m} \left(\exp(ax^m + b) - y^m \right)^2$$

We have to compute a, b to minimize loss.

$$\therefore \frac{\partial L}{\partial a} = \frac{2}{M} \sum_{m} \left(\exp(ax^m + b) - y^m \right) \exp(ax^m + b) \cdot x^m$$

$$\frac{\partial L}{\partial b} = \frac{2}{M} \sum_{m} \left(\exp(ax^m + b) - y^m \right) \exp(ax^m + b)$$

Iteratively: (simultaneously)

$$a_{t+1} = a_t - \alpha \frac{\partial L}{\partial a}$$

$$b_{t+1} = b_t - \alpha \frac{\partial L}{\partial b}$$

Until, $\frac{\partial L}{\partial a} \approx 0$ and $\frac{\partial L}{\partial b} \approx 0$.

Warm-up

Q2. Compute subgradients

(a) $f(x) = \max\{x^2 - 2x, |x|\}$

(i) $x = 0$

$x^2 - 2x = |x| = 0$

gradient does not exist

case1 let $x = -\epsilon$ where ϵ is a small +ve value.

$\therefore f(x) = \max\{\epsilon^2 + 2\epsilon, |-\epsilon|\}$
$= \max\{\epsilon^2 + 2\epsilon, \epsilon\}$

$f(\epsilon) = \epsilon^2 + 2\epsilon$

substituting x,

$f(x) = x^2 - 2x$

$\boxed{\frac{\partial f(x)}{\partial x} = 2x - 2}$ for $x = -\epsilon$ small -ve value

(1)

case2 let $x = \epsilon$ where ϵ is a small +ve value $\& \epsilon < 1$

$\therefore f(x) = \max\{\epsilon^2 - 2\epsilon, \epsilon\}$

Here, $\epsilon^2 - 2\epsilon < \epsilon$
because $\epsilon < 1$

$\therefore\ f(\epsilon) = \epsilon$ for $x=\epsilon$ small +ve value

$\therefore\ \dfrac{\partial f(x)}{\partial x} = 1$ ——— (2)

$\therefore$ From (1) and (2),

Substituting $x=0$ gives us the range of subgradients.

$$2x-2 \leq \left(\frac{\partial f(x)}{\partial x}\right)_{x=0} \leq 1$$

$$-2 \leq \left(\frac{\partial f(x)}{\partial x}\right)_{x=0} \leq 1$$

All the lines who has slope between these slopes are subgradients at $x=0$.

$\therefore\ \dfrac{\partial f(x)}{\partial x} = 0$, is one of the subgradients

$$\boxed{\frac{\partial f(x)}{\partial x} = 0}$$

(ii) $f(x) = \max\{x^2 - 2x, |x|\}$

$x = -2$

$x^2 - 2x = +4 + 4 = 8$

$|x| = 2$

$\therefore f(x) = x^2 - 2x$ $\quad (x^2 - 2x > |x|$ at $x = -2)$

$$\frac{\partial f(x)}{\partial x} = 2x - 2$$

for $x = -2$;

$$\frac{\partial f(x)}{\partial x} = 2(-2) - 2 = -6$$

//

b). $g(x) = \max\{(x-1)^2, (x-2)^2\}$

(i) $x = 1.5$

At $x = 1.5$,
$(x-1)^2 = (x-2)^2 = (0.5)^2 = 0.25$

The gradient does not exist.

Take ϵ = some small +ve value.

Case 1

If $x = 1.5 - \epsilon$

$(x-1)^2 = (0.5-\epsilon)^2$

$(x-2)^2 = (-0.5-\epsilon)^2$

$= (0.5+\epsilon)^2$

$\therefore (x-2)^2 > (x-1)^2$ for $x = 1.5-\epsilon$

$\therefore g(x) = (x-2)^2$

$\therefore \dfrac{\partial g(x)}{\partial x} = 2(x-2).1$

$= 2(x-2)$ —— (1)

Case 2

$x = 1.5+\epsilon$

$(x-1)^2 = (0.5+\epsilon)^2$

$(x-2)^2 = (-0.5+\epsilon)^2$

$= (0.5-\epsilon)^2$

$\therefore (x-1)^2 > (x-2)^2$ for $x = 1.5+\epsilon$

$\therefore g(x) = (x-1)^2$

$\therefore \dfrac{\partial g(x)}{\partial x} = 2(x-1)$

At $x=1.5$,

Substituting $x=1.5$ in ① and ②,

$$2(x-2) \leq \left(\frac{\partial g(x)}{\partial x}\right)_{x=1.5} \leq 2(x-1)$$

$$-1 \leq \frac{\partial g(x)}{\partial x} \leq 1$$

All the lines with this slope at $x=1.5$ are subgradients.

$\frac{\partial g(x)}{\partial(x)} = 0$ is a subgradient at $x=1.5$

(ii) for $g(x) = \max\{(x-1)^2, (x-2)^2\}$

at $x=0$

$g(x) = (x-2)^2$ $\left(\because (x-2)^2 > (x-1)^2 \text{ at } x=0\right)$

$\therefore \frac{\partial g(x)}{\partial x} = 2(x-2)$

$\therefore \boxed{\frac{\partial g(x)}{\partial x} = -4}$

is subgradient at $x=0$ //