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

case 1: 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)

case 2: 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 $\epsilon < 1$

$\therefore \quad f(\epsilon) = \epsilon \qquad$ 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

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

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

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

$x = -2$

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

$|x| = 2$

$\therefore f(x) = x^2 - 2x$ $\quad$ ($x^2 - 2x > |x|$ 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 $x = 1.5 - \epsilon$

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

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

$= 2(x-2)$ —— (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 $x = 1.5 + \epsilon$

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

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

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 $x=0$

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

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

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

is subgradient at $x=0$ //

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.

$$b + -2w_1 + w_2 > 0 \qquad b - 3w_2 < 0$$
$$b + 2w_1 - w_2 > 0 \qquad b + 3w_2 < 0$$

(+) $\qquad\qquad$ (+)

$$2b > 0 \quad -(1) \qquad\qquad 2b < 0 \quad -(2)$$

From (1) and (2);

$$\boxed{b = 0.}$$

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

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

$w_2 > 2w_1$ & $w_1 > 2w_2$

cannot be satisfied with any values of w_1, w_2
$\therefore$ linear seperator does not exist.

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

x_1	x_2	x_1^2	x_2^2	$x_1 x_2$	labels
1	1	1	1	1	+
-1	-1	1	1	1	+
1	-1	1	1	-1	−
-1	1	1	1	-1	−

To find a solution,

$$w_1 x_1^2 + w_2 x_2^2 + w_3 x_1 x_2 + b = 0$$

s.t.

$$w_1 + w_2 + w_3 + b \geq 0 \quad \text{— (1)}$$
$$w_1 + w_2 - w_3 + b < 0 \quad \text{— (2)}$$

The solution to this is

$$\boxed{x_1 x_2 = 0}$$

this is our linear seperator

(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

$$w_1 e^{-1} + w_2 e^{-1} + b > 0 \quad \text{--- (1)}$$

$$w_1 e + w_2 e + b > 0 \quad \text{--- (2)}$$

$$w_1 e^{-1} + w_2 e + b < 0 \quad \text{--- (3)}$$

$$w_1 e + w_2 e^{-1} + b < 0 \quad \text{--- (4)}$$

Add (1) and (2);

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

Add (3) and (4),

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

From (5) and (6),

We conclude that a linear seperator is not possible.

(d) $\phi(x_1, x_2) = \begin{bmatrix} x_1 \sin(x_2) \\ x_1 \end{bmatrix}$

We know that $\sin(-x) = -\sin x$

and let $\sin(1) = c$

x_1	x_2	$x_1 \sin(x_2)$	x_1	labels
-1	-1	c	-1	+
1	1	c	1	+
-1	1	-c	-1	-
1	-1	-c	1	-

$\therefore$ Solution is: $w_1 x_1 \sin(x_2) + w_2 x_1 + b = 0$

s.t.

$$cw_1 - w_2 + b > 0 \qquad -cw_1 - w_2 + b < 0$$
$$cw_1 + w_2 + b > 0 \qquad -cw_1 + w_2 + b < 0$$

$$2cw_1 + 2b > 0 \qquad -2cw_1 + 2b < 0$$

From the table we can see that,

for $x_1 \sin(x_2) > 0$, +ve and

for $x_1 \sin(x_2) < 0$, -ve

$\therefore$ linear separator exists,

given by $\underline{x_1 \sin(x_2) = 0}$

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:

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

In vector notation;

$$f(x) = w^T X + b$$

where $X = \begin{bmatrix} x^K \\ x^{K-1} \\ \vdots \\ x \end{bmatrix}$ $w = \begin{bmatrix} w^K \\ w^{K-1} \\ \vdots \\ w_1 \end{bmatrix}$

Feature vectors

Feature vector X is generated from our input x.

The loss function: mean square error

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

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

And the objective function is given by

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

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

∴ Gradient descent

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

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

Iteratively compute;

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

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

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.

∴ Time Complexity $= O(Km) + O(Km) = O(Km)$

Problem 3 Exponential Regression.

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

M input data points.

Hypothesis

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

loss

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

objective

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

We minimize this function using gradient descent.

(2) 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 $\frac{\partial L}{\partial b} \approx 0$.

(3) Is the optimization problem convex?

Given the loss fu^n, if we can get the convexity for one input, we can say it remains convex by sum of convex fu^n theorem

$$\therefore \quad f(x) = (\exp(ax^i+b) - y^i)^2$$

$$f(x) = (\exp(ax^i+b))^2 - 2\exp(ax^i+b)\, y^i + y^{i^2} \quad \text{①}$$

For the term exponent:

$$(\exp(ax^i+b))^2 = \exp(2(ax^i+b))$$

$$= \underbrace{\exp^{2ax^i}}_{(i)} \cdot \underbrace{\exp^{2b}}_{(ii)}$$

$$\text{(i)} \quad e^{2ax^i} = 2x^i\left(1+a+\frac{a^2}{2!}+\frac{a^3}{3!}\cdots\right)$$

$$\text{(ii)} \quad e^{2b} = 2\left(1+b+\frac{b^2}{2!}+\frac{b^3}{3!}\cdots\right)$$

$\therefore$ $e^{2ax^i} \times e^{2b}$ contains the term $ab, ab^2, a^2b^2 \ldots$

Thus, the eq^n is not convex.

Thus, our loss for exponential regression is not convex.