Homework Assignment 6.

Q1. $f(x) = x^3 + 6x^2 - 3x - 5$

$f'(x) = 3x^2 + 12x - 3$

Solve $f'(x) = 0$.

$3x^2 + 12x - 3 = 0$

Dividing both sides by 3

$x^2 + 4x - 1 = 0$.

using the quadratic formula:-

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

Here, $a = 1$, $b = 1$ and $c = -1$.

$$x = \frac{-4 \pm \sqrt{4^2 - 4(1)(-1)}}{2(1)}.$$

$$x = \frac{-4 \pm \sqrt{16+4}}{2}$$

$$= \frac{-4 \pm \sqrt{20}}{2}$$

$$= \frac{-4 \pm 2\sqrt{5}}{2}$$

$$x = -2 \pm \sqrt{5}$$

So, the stationary points are:-

$x_1 = -2 + \sqrt{5}$ and $x_2 = -2 - \sqrt{5}$

$$f''(x) = 6x+12$$

$$f''(x_1) = 6(-2+\sqrt{5})+12$$

$$= -12+6\sqrt{5}+12$$

$$= 6\sqrt{5}$$

$$f''(x_2) = 6(-2-\sqrt{5})+12$$

$$= -12-6\sqrt{5}+12$$

$$= -6\sqrt{5}$$

Since $-6\sqrt{5} < 0$, $f''(x_2) < 0$.

Thus $x_2 = -2-\sqrt{5}$ is a local maximum

$x_1 = -2+\sqrt{5}$ is a local minimum

7.2 Mini-batch size of one is equal to stochastic gradient descent. Hence, there is no change in the equation.

So,

$$\theta_{i+1} = \theta_i - \gamma_i (\nabla L(\theta_i))^T$$

Stochastic Equation: In Stochastic Gradient Descent (SGD), we consider just one example at a time to take a single step, i.e to update the weights. So,

$$\theta_{i+1} = \theta_i - \gamma_i (\nabla L(\theta_i))^T$$

Mini-Batch Gradient Descent (Mini-Batch = n): Mini-Batch Gradient Descent is similar to Stochastic Gradient descent but instead of considering one example at a time to take a single step it considers n examples to take a single step, i.e to update the weights.

Hence, Mini-Batch Gradient Descent with mini-batch size of one is same as stochastic equation.

7.3.

a) TRUE:-

The intersection of any two convex sets is also a convex. This is because if we take any two points in the intersection of the two convex sets, we can draw a straight line connecting them. Since the points are in intersection of the two convex sets, they are also in each of the convex sets individually. Therefore, the entire line connecting them is also in each of the sets, and the intersection is also convex.

b) FALSE:-

The union of any two convex sets is not necessarily convex. For example, consider two overlapping disks in the plane. Each disk is convex, but their union is not because the boundary of the union is not a convex set.

c) FALSE:-

The difference of a convex set A from another convex set B is not necessarily convex. For example, let A be a unit ball centered at the origin in two dimensions, and let B be a unit ball centered at (1, 0). Then A−B is not convex because it contains a line segment connecting (−1, 0) and (1,0) that is not entirely contained within A−B.

7.4

a) TRUE

Proof:- $f_1(cx_1 + (1-c)x_2) \leq c f_1(x_1) + (1-c) f_1(x_2)$

$f_2(cx_1 + (1-c)x_2) \leq c f_2(x_1) + (1-c) f_2(x_2)$

Then $f_1(cx_1 + (1-c)x_2) + f_2(cx_1 + (1-c) \leq$
$c(f_1(x_1) + f_2(x_1)) + (1-c)(f_1(x_2) + f_2(x_2))$

i.e $f(cx_1 + (1-c)x_2) \leq c f(x_1) + (1-c) x_2$

where $f = f_1 + f_2$

classmate

Date
Page

b) FALSE

For any real-valued function h, $\alpha \in [0,1]$ and x, y in the (convex) domain, let

$D(h, \alpha, x, y) = \alpha h(x) + (1-\alpha) h(y) - h[\alpha x + (1-\alpha) y]$

Convexity for h means $D(h, \alpha, x, y) \geq 0$ for all α, x, y

for this situation, one sufficient condition for $f-g$ to be convex is that $D(f, \alpha, x, y) \geq D(g, \alpha, x, y)$ — ① $\forall \alpha, x, y$

eq ① does not impose convexity directly on f and g.

c) FALSE

The functions $f(x) = 1-x$ and $g(x) = 1+x$ are convex.
However their product $(f*g)(x) = 1-x^2$ is not.

d) TRUE

$f(\lambda x + (1-\lambda)(y)) = \max(f_1(\lambda x + (1-\lambda) y) \leq f_2(\lambda x + (1-\lambda) y)$

$\max(\lambda f_1(x) + (1-\lambda) f_1(y)); \lambda f_2(x) + (1-\lambda) f_2(y) = \lambda$

$\max(f_1(x), f_2(x) + (1-\lambda)$

$\max(f_1(y), f_2(y) = \lambda f(x) + (1-\lambda) f(y).$

7.5 Given:

$$\max_{x \in \mathbb{R}^2, \xi \in \mathbb{R}} p^T x + \xi$$

subject to the constraints that $\xi \geq 0$, $x_0 \leq 0$ and $x_1 \leq 3$.

→

Suppose $P = \begin{bmatrix} p \\ 1 \end{bmatrix}$ and $X = \begin{bmatrix} x \\ \xi \end{bmatrix}$

Then the above problem become:

$$\max_{x \in \mathbb{R}^2} P^T X$$

Subject to the constraints $AX \leq b$

$$A = \begin{bmatrix} 0 & 0 & -1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \text{ and } b = \begin{bmatrix} 0 \\ 0 \\ 3 \end{bmatrix}$$

Hence a standard linear program in matrix notation of the given linear problem is:

$$\max_{X \in \mathbb{R}^2} P^T X$$

Subject to the constraints $AX \leq b$