Tutor: Redion Xhepa
Priyanka Upadhyay, Email Id : s8prupad@stud.uni-saarland.de
Gopal Bhattrai, Email Id : gobh00001@stud.uni-saarland.de

1.a $L = \sum_{j=1}^{K} \sum_{x_i \in S_j} \|x_i - \mu_j\|_2^2$

Since we need to find μ_1 i.e. Only for Cluster 1.

So $L = \sum_{x_i \in S_1} \|x_i - \mu_1\|_2^2$

$$\frac{\partial L}{\partial \mu_1} = \sum_{x_i \in S_1} \frac{\partial}{\partial \mu_1} (x_i - \mu_1)^T (x_i - \mu_1)$$

$$\frac{\partial L}{\partial \mu_1} = \sum_{x_i \in S_1} \frac{\partial}{\partial \mu_1} \left(x_i^T x_i - x_i^T \mu_1 - \mu_1^T x_i + \mu_1^T \mu_1 \right)$$

$$= \sum_{x_i \in S_1} (-2x_i + 2\mu_1)$$

$$= -2 \sum_{x_i \in S_1} (x_i - \mu_1)$$

So.

$$GD \Rightarrow \boxed{\mu_1 = \mu_1 + 2\epsilon \cdot \sum_{x_i \in S_1} (x_i - \mu_1)}$$

1.b SGD is done on a minibatch of Size $p < n$
here $p = 1$, only Instance in our batch x_i

So

$$SGD \Rightarrow \boxed{\mu_1 = \mu_1 + 2\epsilon \cdot (x_i - \mu_1)}$$

1.c In Kmeans the update equation for μ_1 will be :

$$\left[\mu_1 = \frac{1}{n_1} \sum_{x \in S_1} x \right]$$

S_1 contains all points that were assigned Cluster 1

n_1 No. of Points in S_1.

According to Gradient descent ⇒

$$\mu_1 = \mu_1 + 2\epsilon \sum_{x \in S_1} (x - \mu_1)$$

Equating →

$$\frac{1}{n_1} \sum_{x \in S_1} x = \mu_1 + 2\epsilon \sum_{x \in S_1} (x - \mu_1)$$

$$\therefore \left[\epsilon = \frac{\sum_{x \in S_1} \frac{x}{n_1} - \mu_1}{2 \sum_{x \in S_1} (x - \mu_1)} \right]$$

Problem-3 : Hessian and optimization

(a) 1→ Positive, since attains local minimum
2→ Negative, since attains local maximum
3→ Positive, since attains local minimum
4→ Both Positive and Negative, since the curve has a saddle point.

(b) 1→ Positive Definite
2→ Negative Definite
3→ Positive Definite
4→ Neither

Problem - 4 Gradient Descent and Newton's Method $f: R^2 \to R$

→ Minimize the function $f(x) = x_1^2 - 3x_1 + x_2^2 - x_1 x_2$

→ Starting point $x^{(0)} = [1, 1]^T$

$\varepsilon = 0.5$, iteration stop if $L2 < 0.2$

(4.a) 1st Iteration: $x_1 = 1, x_2 = 1$ from starting point

$x_1 = x_1 - 0.5(2x_1 - 3 - x_2)$	$x_2 = x_2 - 0.5(2x_2 - x_1)$
$= 1 - 0.5(2.1 - 3 - 1)$	$= 1 - 0.5(2 - 1)$
$x_1 = 2$	$x_2 = 0.5$

2nd Iteration: $x_1 = 2, x_2 = 0.5$ as initial value

$x_1 = x_1 - 0.5(2x_1 - 3 - x_2)$	$x_2 = x_2 - 0.5(2x_2 - x_1)$
$= 2 - 0.5(2 \times 2 - 3 - 0.5)$	$= 0.5 - 0.5(1 - 2)$
$x_1 = 1.75$	$x_2 = 1$

3rd

$x_1 = 1.75, x_2 = 1$

$x_1 = 1.75 - 0.5(2 \times 1.75 - 3 - 1)$	$x_2 = 1 - 0.5(2 - 1.75)$
$= 2$	$x_2 = 0.875$

4th:

$x_1 = 2, x_2 = 0.875$

$x_1 = 2 - 0.5(4 - 3 - 0.875)$	$x_2 = 0.875 - 0.5(2 \times 0.875 - 2)$
$x_1 = 1.9375$	$x_2 = 1$

(4.b) $f(x) = x_1^2 - 3x_1 + x_2^2 - x_1x_2$

$$g = \begin{bmatrix} 2x_1 - 3 - x_2 \\ 2x_2 - x_1 \end{bmatrix} \qquad h = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$$

for $x^{[0]}$ $[1, 1]$ $\Rightarrow$

$$g = \begin{bmatrix} 2x1 - 3 - 1 \\ 2x1 - 1 \end{bmatrix} = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$$

$$h = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}, \quad h^{-1} \Rightarrow \begin{bmatrix} 2/3 & 1/3 \\ 1/3 & 2/3 \end{bmatrix}$$

so

$$x = x - h^{-1}g = \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} -1 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

Ans

(4.c) If the second derivative is non Negative for it's entire domain, then Newton's method can be applied to a twice continuous function

$f(x) = 12x^3 + 15x$
$f'(x) = 36x^2 + 15$
$f''(x) = 72$

second derivative is non negative w.r.t x so function is convex and Newton's method is applicable.

5.5.a

$f = \lambda x + (1-\lambda) y$ ✓ because f is a Convex function.

This equation Simply means. A function is Convex iff $\forall a, b \in X$ (vector space), the line segment connecting a and b is also in X. $\therefore \lambda \in [0,1]$, $x, y \in$ Vector Space

$\therefore \quad g(x) = f(Ax+b)$

$\therefore$ Lets find $g(\lambda x + (1-\lambda) y)$

$$= f(A(\lambda x + (1-\lambda) y) + b)$$

$$= f(\lambda Ax + (1-\lambda) Ay + b)$$

$$= f(\lambda(Ax+b) + (1-\lambda)(Ay+b))$$

$\rightarrow$ Just open it. Its the same Equation

$$\leq \lambda f(Ax+b) + (1-\lambda) f(Ay+b)$$

because λ is Positive, So ofcourse this quantity will be bigger (This is the Property of being a Convex function)

$= \lambda g(x) + (1-\lambda) g(y) \rightarrow$ Form Resembles

Hence $g(x)$ is also convex.

A function is Convex iff:

$$\lambda f(x) + (1-\lambda) f(x') \geq f(\lambda x + (1-\lambda) x')$$