# DS 122 Homework 10

**Xiang Fu**
xfu@bu.edu
Boston University Faculty of Computing & Data Sciences

# Contents

# 1 Q1 - Gradient Computations

10 Points

## 1.1 Q1.1

Compute the gradient of the function $f(x, y, z) = 3\mathrm{x}^2 y - 2xz^2 + 4y^2 z$.

## 1.2 Solution

Partial Derivative with respect to $x$:

$$\begin{aligned}\frac{\partial f}{\partial x} &= \frac{\partial}{\partial x}(3\mathrm{x}^2 y - 2xz^2 + 4y^2 z) \\ &= 6xy - 2z^2\end{aligned}$$

Partial Derivative with respect to $y$:

$$\begin{aligned}\frac{\partial f}{\partial y} &= \frac{\partial}{\partial y}(3\mathrm{x}^2 y - 2xz^2 + 4y^2 z) \\ &= 3x^2 + 8yz\end{aligned}$$

Partial Derivative with respect to $z$:

$$\begin{aligned}\frac{\partial f}{\partial z} &= \frac{\partial}{\partial z}(3\mathrm{x}^2 y - 2xz^2 + 4y^2 z) \\ &= 4y^2 - 4az\end{aligned}$$

Our gradient is:

$$\nabla f = (6xy - 2z^2, 3x^2 + 8yz, 4y^2 - 4az)$$

## 1.3 Q1.2

Compute the gradient of the function $f(x, y) = e^{xy} + x^2 y^2$.

## 1.4 Solution

Partial Derivative with respect to $x$:

$$\frac{\partial f}{\partial x} = \frac{\partial}{\partial x}(e^{xy} + x^2y^2)$$
$$= e^{xy} \cdot y + 2xy^2$$

Partial Derivative with respect to $y$:

$$\frac{\partial f}{\partial y} = \frac{\partial}{\partial y}(e^{xy} + x^2y^2)$$
$$= e^{xy} \cdot x + 2x^2y$$

Our gradient is:

$$\nabla f = \left(e^{xy} \cdot y + 2xy^2, e^{xy} \cdot x + 2x^2y\right)$$

## 1.5 Q1.3

Let $f(x, y) = \sin(x)\cos(y)$. Find the gradient of $f$ at point $\left(\frac{\pi}{4}, \frac{\pi}{4}\right)$.

## 1.6 Solution

Partial Derivative with respect to $x$:

$$\frac{\partial f}{\partial x} = \frac{\partial}{\partial x}(\sin(x)\cos(y))$$
$$= \cos(x)\cos(y)$$

Partial Derivative with respect to $y$:

$$\frac{\partial f}{\partial y} = \frac{\partial}{\partial y}((\sin(x)\cos(y))$$
$$= -\sin(x)\sin(y)$$

Evaluate at $\left(\frac{\pi}{4}, \frac{\pi}{4}\right)$.

$$\left(\frac{\partial f}{\partial x}\right)_{|\frac{\pi}{4},\frac{\pi}{4}} = \cos\left(\frac{\pi}{4}\right)\cos\left(\frac{\pi}{4}\right) = \frac{\sqrt{2}}{2} * \frac{\sqrt{2}}{2} = \frac{1}{2}$$

$$\left(\frac{\partial f}{\partial y}\right)_{|\frac{\pi}{4},\frac{\pi}{4}} = -\sin\left(\frac{\pi}{4}\right)\sin\left(\frac{\pi}{4}\right) = -\frac{\sqrt{2}}{2} * \frac{\sqrt{2}}{2} = -\frac{1}{2}$$

So, our gradient of $f$ at point $\left(\frac{\pi}{4}, \frac{\pi}{4}\right)$ is $\left(\frac{1}{2}, -\frac{1}{2}\right)$.

# 2 Q2 - Quadratic Forms

## 2.1 Q2.1

Recall that every quadratic form can be expressed as $x^T Ax$ where A is a symmetric matrix. What quadratic form corresponds to the matrix

$$\begin{Bmatrix} 4 & 1 \\ 1 & 2 \end{Bmatrix}$$

## 2.2 Solution

Multiply $x^T$ with A:

$$\{x \;\; y\} * \begin{Bmatrix} 4 & 1 \\ 1 & 2 \end{Bmatrix} = \{4x + y \;\; x + 2y\}$$

Multiply this result by $x$:

$$\{4x + y \;\; x + 2y\} * \begin{Bmatrix} x \\ y \end{Bmatrix}$$

$$= (4x + y) \cdot x + (x + 2y) \cdot y$$

$$= 4x^2 + xy + xy + 2y^2$$

$$= 4x^2 + 2xy + 2y^2$$

$$x^T Ax = 4x^2 + 2xy + 2y^2$$

## 2.3 Q2.2

What is the symmetric matrix A that corresponds to the quadratic form $4x_1^2 + 2x_1x_2 + x_2^2$?

## 2.4 Solution

Vector $x = \begin{Bmatrix} x_1 \\ x_2 \end{Bmatrix}$

The quadratic form can be rewritten as:

$$x_1^2 \cdot 4 + x_1 x_2 \cdot 2 + x_2^2 \cdot 1$$

The coefficients in the quadratic form correspond to the entries in the matrix $A$. Since $A$ is symmetric, it will have the form:

$$A = \begin{Bmatrix} a & b \\ b & c \end{Bmatrix}$$

Matching the coefficients:

$a$ corresponds to the coefficient of $x_1^2$ = 4

$c$ corresponds to the coefficient of $x_2^2$ = 1

Because in the expansion $x^T Ax$, each off-diagonal term appears twice. Therefore, $b$ corresponds to half the coefficient of $x_1 x_2$, which is $b = \frac{2}{2} = 1$.

Matrix $A =$

$$\begin{Bmatrix} 4 & 1 \\ 1 & 1 \end{Bmatrix}$$

# 3 Q3 - Gradient Descent

## 3.1 Q3.1

Perform the first two steps of gradient descent on the function $f(x) = x^4 - 6x^3 + 12x^2 - 8x$ with an initial guess of $x$ = 2 and a learning rate of 0.01. Write out all your steps.

## 3.2 Solution

Given:

$f(x) = x^4 - 6x^3 + 12x^2 - 8x$

Initial guess $x_0 = 2$

$f'(x) = 4x^3 - 18x^2 + 24x - 8$

So,

$$x_{\text{new}} = x_{\text{odd}} - \alpha \cdot f'(x_{\text{odd}})$$

**Firt Iteration**

$$f'(2) = 4(2)^3 - 18(2)^2 + 24(2) - 8 = 0$$

$$x_1 = 2 - 0.01 \cdot 0 = 2$$

And as we can see here, since the derivative at x = 2 is zero, the update does not change the value of x.

**Second Iteration**

Since $x_1 = 2$ and the derivative at this point is 0, the value of x will not change in the second iteration.

$$\therefore x_2 = x_1 = 0$$

After two iterations of gradient descent, the value of x remains at 2. This is because the derivative of the function at x = 2 is zero, indicating that this point is either a local minimum, maximum, or saddle point. And so the algorithm has converged to a stationary point.

## 3.3 Q3.2

Perform the first two steps of gradient descent on the function $f(x) = 3x^2 + 5y^2$ with an initial guess of $x = 1$ and $y = 1$ and a learning rate of 0.05. Write out all your steps.

## 3.4 Solution

Find the gradient:

$$\nabla f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right) = (6x, 10y)$$

We start with the initial guess at $x_0 = 1$ and $y_0 = 1$.

First Iteration

Compute the gradient at $(1, 1)$:

$$\nabla f(1, 1) = (6 \cdot 1, 10 \cdot 1) = (6, 10)$$

Update $x$ and $y$

$$x_1 = 1 - 0.05 \cdot 6 = 1 - 0.3 = 0.7$$

$$y_1 = 1 - 0.05 \cdot 10 = 1 - 0.5 = 0.5$$

Second Iteration

Compute the gradient at $(0.7, 0.5)$:

$$\nabla f(0.7, 0.5) = (6 \cdot 0.7, 10 \cdot 0.5) = (4.2, 5)$$

Update $x$ and $y$

$$x_2 = 0.7 - 0.05 \cdot 4.2 = 0.7 - 0.21 = 0.49$$

$$y_2 = 0.5 - 0.05 \cdot 5 = 0.5 - 0.25 = 0.25$$

So after two iterations of gradient descent, the value of $x_2$ and $y_2$ changed to 0.49 and 0.25. It shows that gradient algorithm is moviong towards the minimum of the function $f(x, y) = 3x^2 + 5y^2$

## 3.5 Q3.3

You will perform the first step of a linear regression by gradient descent. Your loss function is defined as $L(w) = ||y - XB||^2$ and your initial guess for the slope $B_1$ is 1 and for the intercept $B_0$ is 0.

There are only three data points for your regression, $x = (10, 15, 20)$ with corresponding $y = (10, 20, 30)$. For a learning rate of $0.05$, what will be your next setting of slope $B_1$ and intercept $B_0$? Write out all your steps.

## 3.6 Solution

Setup

$$X = \begin{Bmatrix} 1 & 10 \\ 1 & 15 \\ 1 & 20 \end{Bmatrix}$$

$$y = \begin{Bmatrix} 1 \\ 20 \\ 30 \end{Bmatrix}$$

$$B = \begin{Bmatrix} B_0 \\ B_1 \end{Bmatrix} = \begin{Bmatrix} 0 \\ 1 \end{Bmatrix}$$

Compute the predictions

$$\hat{y} = XB = \begin{Bmatrix} 1 & 10 \\ 1 & 15 \\ 1 & 20 \end{Bmatrix} \cdot \begin{Bmatrix} 0 \\ 1 \end{Bmatrix} = \begin{Bmatrix} 10 \\ 15 \\ 20 \end{Bmatrix}$$

Compute the loss gradient

The loss function as given, is $L(w) = ||y - XB||^2$

The gradient of the loss with respect to $B$ is $\nabla_B L = -2X^T(y - \hat{y})$

Compute $y - \hat{y}$

$$\begin{Bmatrix} 10 \\ 20 \\ 30 \end{Bmatrix} - \begin{Bmatrix} 10 \\ 15 \\ 20 \end{Bmatrix} = \begin{Bmatrix} 0 \\ 5 \\ 10 \end{Bmatrix}$$

Compute $\nabla_B L$

$$-2 \cdot \begin{Bmatrix} 1 & 1 & 1 \\ 10 & 15 & 20 \end{Bmatrix} \cdot \begin{Bmatrix} 0 \\ 5 \\ 10 \end{Bmatrix}$$

$$= -2 \cdot \begin{Bmatrix} 15 \\ 250 \end{Bmatrix}$$

Update the coefficients

Given the learning rate $\alpha = 0.05$

$$B_0 = B_0 - \alpha \cdot \nabla_{B_0} L = 0 - 0.05 \cdot (-2 \cdot 15) = 0.05 \cdot 30 = 1.5$$

$$B_1 = B_1 - \alpha \cdot \nabla_{B_1} L = 1 - 0.05 \cdot (-2 \cdot 250) = 1 + 0.05 \cdot 500 = 26$$

So, after the first iteration of gradient descent, the updated slope $B_1$ is 26 and the intercept $B_0$ is 1.5.

# 4 Q4 Gradient Descent With Momentum

Calculate the first two steps of gradient descent with momentum for the function $f(x) = x^3 - 3x$ with an initial guess of $x = 1$, a learning rate of 0.05, and a momentum factor of 0.1. Write out all your steps.

## 4.1 Solution

Given:

- Our function: $f(x) = x^3 - 3x$
- Initial guess: $x_0 = 1$
- Initial velocity: $v_0 = 0$
- Learning rate: $\alpha = 0.05$

- Momentum factor: $\beta = 0.1$

The gradient of $f(x) = x^3 - 3x$ is

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

**First Iteration**

Compute the derivative at $x_0$

$$x_0 = f'(1) = 3 \cdot (1)^2 - 3 = 0$$

Update the velocity

$$v_1 = \beta v_0 - \alpha f'(1) = 0.1 \cdot 0 - 0.05 \cdot 0 = 0$$

Update the postion

$$x_1 = x_0 + v_1 = 1 + 0 = 1$$

**Second Iteration**

Compute the derivative at $x_1$

$$x_1 = f'(1) = 3 \cdot (1)^2 - 3 = 0$$

Update the velocity

$$v_2 = \beta v_1 - \alpha f'(1) = 0.1 \cdot 0 - 0.05 \cdot 0 = 0$$

Update the position

$$x_2 = x_1 + v_2 = 1 + 0 = 1$$

Since the derivative of the function at $x = 1$ is 0, indicating that this point is a stationary point, after two iterations of gradient descent with momentum, the value of $x$ still remains at 1. Thus, the momentum does not have any effect in this case since the velocity updates are 0 due to the zero derivative.