Quiz-3

Due Mar 20 at 23:59 **Points** 10 **Questions** 10
Available Mar 19 at 19:30 - Mar 20 at 23:59 **Time Limit** 30 Minutes

Instructions

The quiz is set to open on March 19th at 7:30 PM and closes on Monday, March 20th at 11:59 PM.

Attempt History

	Attempt	Time	Score
LATEST	Attempt 1	28 minutes	4 out of 10

ⓘ Correct answers will be available Mar 25 at 0:00 - Mar 26 at 0:00.

Score for this quiz: **4** out of 10
Submitted Mar 19 at 19:58
This attempt took 28 minutes.

Incorrect

Question 1 — 0 / 1 pts

Choose the incorrect option about the momentum based learning.

- (●) Useful when gradient needs to be approximated.
- ( ) None of the other options is true
- ( ) Solves issues of local optima and flat regions.
- ( ) Uses the feedback from previous term

Question 2 — 1 / 1 pts

In case of momentum-based learning using the formula, $v \leftarrow \beta v - \alpha \frac{\partial J}{\partial w}$, where β = 0.9, we can make the following statement:

- None of the other options is true
- In a flat region the length of the update vector will be the same at each update
- In a flat region, the length of the update vector will keep decreasing
- In a flat region the length of the update vector will keep increasing because of existing momentum

Incorrect

Question 3

0 / 1 pts

In case of momentum-based learning using the formula, $v \leftarrow \beta v - \alpha \frac{\partial J}{\partial w}$, where β = 1.0, we can make the following statement:

- None of the other options is true
- In a flat region the length of the update vector will be the same at each update
- In a flat region the length of the update vector will keep increasing because of existing momentum
- In a flat region, the length of the update vector will keep decreasing

Question 4

1 / 1 pts

For any function $f(x, y)$ it is possible to get to the minimum of the function in a single-step of sufficient size starting at any point using gradient descent

- ○ True
- ◉ False

Incorrect

Question 5

0 / 1 pts

Consider data points $(x_i, y_i), i = 1 \ldots\ldots n.$ If variance of x_i is 0.01 and variance of y_i is 0.65. For dimension reduction,

- ◉ Projection of data onto x axis is better than y axis
- ○ Projection of data onto z axis is better
- ○ Projection of data onto either of the axis gives same result.
- ○ Projection of data onto y axis is better than x axis

Incorrect

Question 6

0 / 1 pts

Momentum based gradient descent is faster than stochastic gradient descent

- ○ True
- ◉ False

Question 7

1 / 1 pts

Let $f(x) = x^4 - 3x^3 + 2x^2$. Choose the best initial point from the following for minimization.

- 0.25
- 0.1
- 1.5
- -0.1

Question 8

1 / 1 pts

There are around N data points and each data point belongs to $\mathbf{R}^n$. Dimension reduction means reduction of

- n
- None of the other options is true
- N
- Both N and n

Incorrect

Question 9

0 / 1 pts

Consider the case of gradient descent with momentum learning with the formula for the new update vector, $v \leftarrow \beta v - \alpha \frac{\partial J}{\partial w}$, where the current gradient of the loss function with respect to the parameters w_1 and w_2 , i.e $\frac{\partial J}{\partial w}$ is [0.1 0.2] and the previous update vector was [1.5

2.0] . Let the momentum parameter β =0.5 and the learning rate α =0.1. The new update vector will be

- [0.76 1.02]
- [0.74 0.98]
- None of the other options is true
- [0.75 1.0]

Incorrect

Question 10

0 / 1 pts

Consider two data points $x_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $x_2 = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$ and . The variance left over after finding the first principal component is

- Zero
- One
- Infinity
- None of the other options

Quiz Score: **4** out of 10

Quiz-3

Due Mar 20 at 23:59 **Points** 10 **Questions** 10
Available Mar 19 at 19:30 - Mar 20 at 23:59 **Time Limit** 30 Minutes

Instructions

The quiz is set to open on March 19th at 7:30 PM and closes on Monday, March 20th at 11:59 PM.

Attempt History

	Attempt	Time	Score
LATEST	Attempt 1	5 minutes	6 out of 10

Correct answers will be available Mar 25 at 0:00 - Mar 26 at 0:00.

Score for this quiz: **6** out of 10
Submitted Mar 19 at 20:46
This attempt took 5 minutes.

Incorrect

Question 1 — 0 / 1 pts

let $f(x) = x^T x, x \in R^2$ and let $x_0 = [1, 1]^T$. Choose the best first step size for gradient descent from the following

- 0.4
- 0.75
- 0.1
- 0.25

Question 2

1 / 1 pts

Let $f(x) = x^4 - 3x^3 + 2x^2$. Choose the best initial point from the following for minimization.

- (●) 1.5
- ( ) 0.1
- ( ) -0.1
- ( ) 0.25

Question 3

1 / 1 pts

Consider the case of gradient descent with momentum learning with the formula for the new update vector, $v \leftarrow \beta v - \alpha \frac{\partial J}{\partial w}$ where the current gradient of the loss function with respect to the parameters w_1 and w_2 , i.e $\frac{\partial J}{\partial w}$ is [0.1 0.2] and the previous update vector was [1.5 2.0] . Let the momentum parameter β =0.5 and the learning rate α =0.1. The new update vector will be

- ( ) [0.75 1.0]
- ( ) [0.76 1.02]
- (●) [0.74 0.98]
- ( ) None of the other options is true

Question 4

1 / 1 pts

Which of the statements below is true with respect to gradient descent on a function $L(x, y)$ at the point (x_0, y_0) ?

- If we find out at the point (x_0, y_0) that $\frac{\partial^2 L}{\partial x^2}$ is very different from $\frac{\partial^2 L}{\partial y^2}$ then we can carry on with the same learning rate for both x and y
- **If we find out at the point (x_0, y_0) that $\frac{\partial^2 L}{\partial x^2}$ is very different from $\frac{\partial^2 L}{\partial y^2}$ then we cannot carry on with the same learning rate for both x and y**
- The second derivatives calculated with respect to x_0 and y_0 give no insight into the value of the learning rate at (x_0, y_0) .
- None of the other options is true

Question 5

1 / 1 pts

Gradient Descent is an optimization algorithm used for

- maximizing the cost function in various machine learning algorithms
- **minimizing the cost function in various machine learning algorithms**
- remaining same the cost function in various machine learning algorithms
- Certain Changes in algorithm

Question 6

1 / 1 pts

Consider the dataset given below, and let the loss function J be the least-squared regression loss function, i.e sum of squared differences. Which of the statements below is true?

X_1	X_2	X_3
1	0	0
0	1	1

- $\frac{\partial J}{(\partial w_1)} = w_1 \; and \; \frac{\partial J}{(\partial w_2)} = 2w_2 - 2.$
- None of the other options is true
- **$\frac{\partial J}{(\partial w_1)} = 2w_1 \; and \; \frac{\partial J}{(\partial w_2)} = 2w_2 - 2.$**
- $\frac{\partial J}{(\partial w_1)} = 2w_1 \; and \; \frac{\partial J}{(\partial w_2)} = 2w_2$

Incorrect

Question 7

0 / 1 pts

If the data covariance matrix $S = \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}$, then

- **A line passing through origin other than X and Y axis is the principal subspace of dim 1.**
- None of the other options is true
- X axis is the principal subspace of dim 1

Y axis is the principal subspace of dim 1

Incorrect

Question 8

0 / 1 pts

For any given dataset the PCA procedure will always obtain a unique set of principal components to represent the data

- (•) True
- ( ) False

Question 9

1 / 1 pts

Regular gradient descent will lead to oscillatory behavior during the minimization of the function $f(x, y) = e^{(-0.01x)}y^2$ in the region $0 \leq x \leq 10, -\infty < y < \infty$

- (•) True
- ( ) False

Incorrect

Question 10

0 / 1 pts

In case of momentum-based learning using the formula, $v \leftarrow \beta v - \alpha \frac{\partial J}{\partial w}$ where β = 0.9, we can make the following statement:

- ( ) In a flat region, the length of the update vector will keep decreasing

In a flat region the length of the update vector will keep increasing because of existing momentum

In a flat region the length of the update vector will be the same at each update

None of the other options is true

Quiz Score: **6** out of 10

Quiz-3

Due Mar 20 at 23:59 **Points** 10 **Questions** 10
Available Mar 19 at 19:30 - Mar 20 at 23:59 **Time Limit** 30 Minutes

Instructions

The quiz is set to open on March 19th at 7:30 PM and closes on Monday, March 20th at 11:59 PM.

Attempt History

	Attempt	Time	Score
LATEST	Attempt 1	26 minutes	8 out of 10

ⓘ Correct answers will be available Mar 25 at 0:00 - Mar 26 at 0:00.

Score for this quiz: **8** out of 10
Submitted Mar 19 at 20:52
This attempt took 26 minutes.

Question 1 — 1 / 1 pts

Momentum based gradient descent is faster than stochastic gradient descent

- (●) True
- ( ) False

Question 2 — 1 / 1 pts

Which of the statements below is true with respect to gradient descent on a function $L(x, y)$ at the point (x_0, y_0) ?

- If we find out at the point (x_0, y_0) that $\frac{\partial^2 L}{\partial x^2}$ is very different from $\frac{\partial^2 L}{\partial y^2}$ then we can carry on with the same learning rate for both x and y
- If we find out at the point (x_0, y_0) that $\frac{\partial^2 L}{\partial x^2}$ is very different from $\frac{\partial^2 L}{\partial y^2}$ then we cannot carry on with the same learning rate for both x and y
- The second derivatives calculated with respect to x_0 and y_0 give no insight into the value of the learning rate at (x_0, y_0) .
- None of the other options is true

Incorrect

Question 3

0 / 1 pts

Let $f(x) = 4x^4 + x^2, g(x) = -3 + 2x$. Then maximize $f(x)$ subject to condition $g(x) \leq 0$. Choose the most appropriate option

- Strong duality holds true for the above primal dual problem.
- all of the other options is true
- f is convex
- g is convex

Question 4

1 / 1 pts

Pick out the number of updates in one epoch for mini batch gradient descent algorithm where N is the number of data points and B is the mini batch size

- (●) N/B
- ( ) 1
- ( ) N
- ( ) B

Question 5

1 / 1 pts

Let $f(x) = x^4 - 3x^3 + 2x^2$. Choose the best initial point from the following for minimization.

- ( ) 0.1
- ( ) -0.1
- ( ) 0.25
- (●) 1.5

Question 6

1 / 1 pts

Consider the function
$G(x_1, x_2, x_3, x_4, x_5 \dots x_n) = f_1(x_1) + f_2(x_2) + \cdots f_n(x_n)$ Let

each function $f_i(x_i)$ have two distinct local minima. Then the number of local minima for $G(x_1, x_2, x_3, x_4, x_5 \dots x_n)$ is

- 2
- 2n
- **2^n**
- None of the other options is true

Incorrect

Question 7

0 / 1 pts

Feature normalization is a technique used to reduce

- Multiple local minima
- All of the other options is true
- The flat surface
- Vanishing and exploding gradient

Question 8

1 / 1 pts

Consider the function $f(x, y) = ax^2 + bxy + cy^2$ where a, b, c are all non-zero. This function has a local minimum within the domain $-1 \leq x \leq 1, -1 \leq y \leq 1.$

- True
- False

Question 9

1 / 1 pts

Consider two data points $x_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $x_2 = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$ and . The variance left over after finding the first principal component is

- One
- None of the other options
- Infinity
- Zero

Question 10

1 / 1 pts

Choose the incorrect option about the momentum based learning.

- Uses the feedback from previous term
- Solves issues of local optima and flat regions.
- Useful when gradient needs to be approximated.
- None of the other options is true

Quiz Score: 8 out of 10

Quiz-3

Due Mar 20 at 23:59 **Points** 10 **Questions** 10
Available Mar 19 at 19:30 - Mar 20 at 23:59 **Time Limit** 30 Minutes

Instructions

The quiz is set to open on March 19th at 7:30 PM and closes on Monday, March 20th at 11:59 PM.

Attempt History

	Attempt	Time	Score
LATEST	Attempt 1	20 minutes	7 out of 10

ⓘ Correct answers will be available Mar 25 at 0:00 - Mar 26 at 0:00.

Score for this quiz: **7** out of 10
Submitted Mar 19 at 21:15
This attempt took 20 minutes.

Question 1 — 1 / 1 pts

Pick out the number of updates in one epoch for mini batch gradient descent algorithm where N is the number of data points and B is the mini batch size

- ○ N
- ○ 1
- ◉ N/B
- ○ B

Incorrect

Question 2 — 0 / 1 pts

Consider data points $(x_i, y_i), i = 1 \ldots\ldots n.$ If variance of x_i is 0.01 and variance of y_i is 0.65. For dimension reduction,

- Projection of data onto x axis is better than y axis
- Projection of data onto z axis is better
- Projection of data onto y axis is better than x axis
- Projection of data onto either of the axis gives same result.

Question 3

1 / 1 pts

Consider the function $G(x_1, x_2, x_3, x_4, x_5 \ldots x_n) = f_1(x_1) + f_2(x_2) + \cdots f_n(x_n)$ Let each function $f_i(x_i)$ have two distinct local minima. Then the number of local minima for $G(x_1, x_2, x_3, x_4, x_5 \ldots x_n)$ is

- 2
- None of the other options is true
- 2n
- 2^n

Question 4

1 / 1 pts

If the data covariance matrix $S = \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}$, then

None of the other options is true

X axis is the principal subspace of dim 1

Y axis is the principal subspace of dim 1

A line passing through origin other than X and Y axis is the principal subspace of dim 1.

Question 5

1 / 1 pts

Consider data points $(x_i, y_i), i = 1 \ldots\ldots n$. If $\sum_{i=1}^{n} x_i^2$ = 0.645 and $\sum_{i=1}^{n} y_i^2$ = 0.06 where (x_i, y_i) are centered at 0. For dimension reduction,

Projection of data onto either of the axis gives same result

Projection of data onto y axis is better than x axis.

Projection of data onto x axis is better than y axis.

Projection of data onto z axis is better

Question 6

1 / 1 pts

Gradient Descent is an optimization algorithm used for

maximizing the cost function in various machine learning algorithms

remaining same the cost function in various machine learning algorithms

Certain Changes in algorithm

minimizing the cost function in various machine learning algorithms

Question 7

1 / 1 pts

Consider the case of gradient descent with momentum learning with the formula for the new update vector, $v \leftarrow \beta v - \alpha \frac{\partial J}{\partial w}$ where the current gradient of the loss function with respect to the parameters w_1 and w_2 , i.e $\frac{\partial J}{\partial w}$ is [0.1 0.2] and the previous update vector was [1.5 2.0] . Let the momentum parameter β =0.5 and the learning rate α =0.1. The new update vector will be

[0.74 0.98]

[0.75 1.0]

[0.76 1.02]

None of the other options is true

Question 8

1 / 1 pts

In case of momentum-based learning using the formula, $v \leftarrow \beta v - \alpha \frac{\partial J}{\partial w}$ where β = 1.0, we can make the following statement:

None of the other options is true

In a flat region, the length of the update vector will keep decreasing

In a flat region the length of the update vector will be the same at each update

In a flat region the length of the update vector will keep increasing because of existing momentum

Incorrect

Question 9

0 / 1 pts

Let $f(x) = 4x^4 + x^2, g(x) = -3 + 2x$. Then maximize $f(x)$ subject to condition $g(x) \leq 0$. Choose the most appropriate option

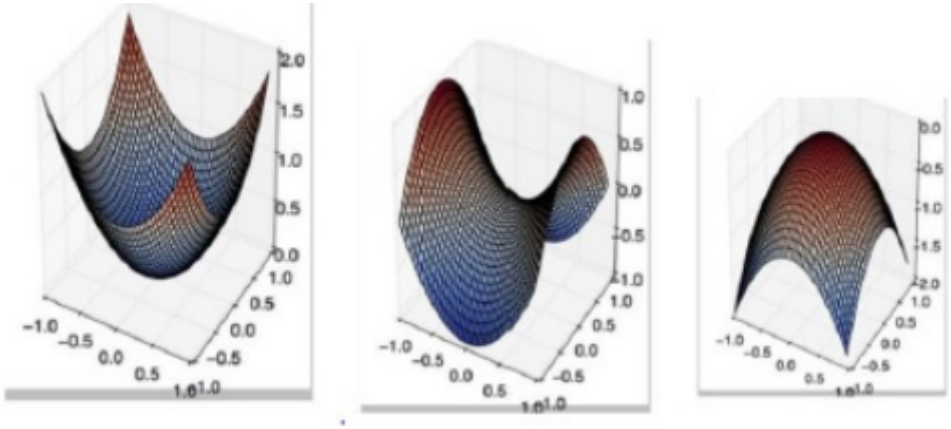
- f is convex
- all of the other options is true
- Strong duality holds true for the above primal dual problem.
- g is convex

Incorrect

Question 10

0 / 1 pts

Can you identify in which of the following graphs Gradient descent will not work properly



- (●) Third figure
- ( ) First and second figures
- ( ) First figure
- ( ) Second figure

Quiz Score: **7** out of 10

Quiz-3 Results for SAIBALA SUNDRAM

ⓘ Correct answers will be available Mar 25 at 0:00 - Mar 26 at 0:00.

Score for this quiz: **7** out of 10
Submitted Mar 19 at 21:16
This attempt took 12 minutes.

Question 1

1 / 1 pts

Consider data points $(x_i, y_i), i = 1 \ldots\ldots n$. If $\sum_{i=1}^{n} x_i^2$ = 0.645 and $\sum_{i=1}^{n} y_i^2$ = 0.06 where (x_i, y_i) are centered at 0. For dimension reduction,

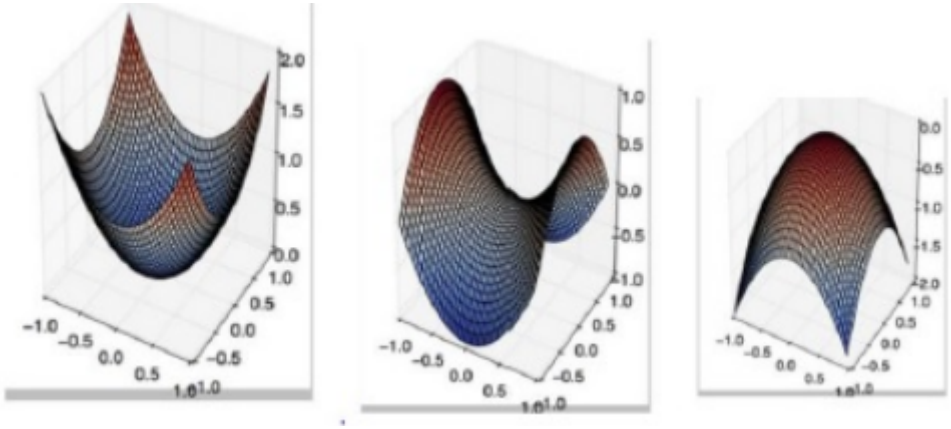
- ○ Projection of data onto either of the axis gives same result
- ○ Projection of data onto z axis is better
- ◉ Projection of data onto x axis is better than y axis.
- ○ Projection of data onto y axis is better than x axis.

Incorrect

Question 2

0 / 1 pts

Can you identify in which of the following graphs Gradient descent will not work properly



- Second figure
- First figure
- First and second figures
- Third figure

Question 3

1 / 1 pts

For any function $f(x, y)$ it is possible to get to the minimum of the function in a single-step of sufficient size starting at any point using gradient descent

- True
- False

Question 4

1 / 1 pts

Gradient Descent is an optimization algorithm used for

- remaining same the cost function in various machine learning algorithms

- maximizing the cost function in various machine learning algorithms
- minimizing the cost function in various machine learning algorithms
- Certain Changes in algorithm

Question 5

1 / 1 pts

Consider the case of gradient descent with momentum learning with the formula for the new update vector, $v \leftarrow \beta v - \alpha \frac{\partial J}{\partial w}$, where the current gradient of the loss function with respect to the parameters w_1 and w_2 , i.e $\frac{\partial J}{\partial w}$ is [0.1 0.2] and the previous update vector was [1.5 2.0] . Let the momentum parameter β =0.5 and the learning rate α =0.1. The new update vector will be

- [0.76 1.02]
- [0.74 0.98]
- None of the other options is true
- [0.75 1.0]

Question 6

1 / 1 pts

Momentum based gradient descent is faster than stochastic gradient descent

- True
- False

Incorrect

Question 7

0 / 1 pts

If the data covariance matrix $S = \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}$, then

- A line passing through origin other than X and Y axis is the principal subspace of dim 1.
- Y axis is the principal subspace of dim 1
- X axis is the principal subspace of dim 1
- None of the other options is true

Question 8

1 / 1 pts

Consider the function $f(x, y) = ax^2 + bxy + cy^2$ where a, b, c are all non-zero. This function has a local minimum within the domain $-1 \leq x \leq 1, -1 \leq y \leq 1$.

- True
- False

Incorrect

Question 9

0 / 1 pts

Which of these statements about mini-batch gradient descent is true?

One iteration of mini-batch gradient descent (computing on a single mini-batch) is faster than one iteration of batch gradient descent

Training one epoch (one pass through the training set) using mini-batch gradient descent is faster than training one epoch using batch gradient descent

You should implement mini-batch gradient descent without an explicit for-loop over different mini-batches, so that the algorithm processes all mini-batches at the same time (vectorization).

None of the other options is true

Question 10

1 / 1 pts

let $f(x) = x^T x, x \in R^2$ and let $x_0 = [1, 1]^T$. Choose the best first step size for gradient descent from the following

0.75

0.4

0.1

0.25

Quiz Score: 7 out of 10