Mini-Assignment: Introduction to Machine Learning

Due Jan 23 at 11:59pm   **Points** 4   **Questions** 2
Time Limit None   **Allowed Attempts** 2

Take the Quiz Again

Attempt History

	Attempt	**Time**	**Score**
LATEST	**Attempt 1**	less than 1 minute	4 out of 4

⚠ Answers will be shown after your last attempt

Score for this attempt: **4** out of 4
Submitted Jan 20 at 6:41pm
This attempt took less than 1 minute.

Question 1	**2 / 2 pts**

Consider two events, A and B. These two events are DISJOINT. If P(B) > 0, what is P(A|B)?

○ 0.5

◉ 0

> Correct, disjoint events cannot co-occur so the probability of A occuring given that B has occurred is zero.

○ 1

○ 0.25

Question 2   **2 / 2 pts**

Suppose that the PDF of a random variable is as follows:

$$f(x) = \begin{array}{ll} \frac{4}{3}(1 - x^3) & \text{for } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{array}$$

What is the value of P(X < 0)?

○ 0.5

○ 2/3

○ 1

◉ 0

Correct, the PDF is 0 whenever X is less than 0, so the total probability is also 0.

Quiz Score: **4** out of 4

Unit 1: Quiz

Due Jan 24 at 11:59pm   **Points** 3   **Questions** 3
Available until Jan 25 at 3am   **Time Limit** None

This quiz was locked Jan 25 at 3am.

Attempt History

	Attempt	Time	Score
LATEST	**Attempt 1**	less than 1 minute	3 out of 3

Score for this quiz: **3** out of 3
Submitted Jan 20 at 6:41pm
This attempt took less than 1 minute.

Question 1   1 / 1 pts

A car deal is organizing his inventory into three types. gas-based, electric, and hybrid. Which data type would most appropriate for this task?

Correct!

- ⦿ Categorical

- ○ Ordinal

- ○ Numerical

Question 2   1 / 1 pts

Which of the following methods would be used to discover groups of like-minded visitors visiting a blog site?

○ Regression

Correct!

◉ Clustering

○ Dimensionality reduction

○ Classification

Question 3   **1 / 1 pts**

Which of the following is a data type that is most useful for representing the data on webpages and the hyperlinks?

Correct!

◉ Graphs with edges and nodes

○ Nominal

○ Ordinal

Quiz Score: **3** out of 3

Unit 2: Quiz

Due Jan 31 at 11:59pm　　　　**Points** 12　　　　**Questions** 12
Available until Feb 1 at 3am　　　　**Time Limit** None

This quiz was locked Feb 1 at 3am.

Attempt History

	Attempt	Time	Score
LATEST	**Attempt 1**	7 minutes	12 out of 12

Score for this quiz: **12** out of 12
Submitted Jan 20 at 6:50pm
This attempt took 7 minutes.

Question 1　　　　　　　　　　　　　　　　1 / 1 pts

Given two sets, A={1, 2, 3, 4, 5}, B={1, 4, 7, 9, 10}. What is A ∩ B?

Correct!

- ⦿ {1, 4}

- ○ {1, 2, 3, 4, 5}

- ○ {1, 2, 3, 4, 5, 7, 9, 10}

- ○ {1, 4, 7, 9, 10}

Question 2　　　　　　　　　　　　　　　　1 / 1 pts

What is $\dfrac{\partial\left(x^{T}ABx\right)}{\partial x}$ if AB is symmetric?

○ x

○ x^T

Correct!   ◉ $2ABx$

○ $2AB$

Question 3   1 / 1 pts

Given matrix $X = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$, what is the value of trace *(X)?*

○ 3

Correct!   ◉ 5

○ 7

○ 4

Question 4   1 / 1 pts

Which of the following is always correct?

○ trace(X+Y)= trace(X)- trace(Y)

○ trace(X+Y)= trace(X)/trace(Y)

Correct!   ◉ trace(X+Y)= trace(X)+trace(Y)

○ trace(X+Y)= trace(X)* trace(Y)

Question 5

1 / 1 pts

The table below shows the purchase history of 10 customers from a set of zip codes that bought organic tea or organic coffee. Using Bayes' Rule, what is the probability that a person who lives in the 44005 zip code and bought organic coffee will likely not buy the organic tea?

CustomerID	Zip Code	Bought Organic Coffee	Bought Organic Tea
1	44005	Yes	Yes
2	44001	No	No
3	44001	Yes	Yes
4	44005	No	No
5	44003	Yes	No
6	44005	No	Yes
7	44005	No	No
8	44001	No	No
9	44005	Yes	Yes
10	44003	Yes	Yes

- ○ 0.24

- ○ 0.5

Correct! ◉ 0

- ○ 0.76

Question 6

1 / 1 pts

The table below shows the purchase history of 10 customers from a set of zip codes that bought organic tea or organic coffee. What is the prior probability that a customer came from area with zip code 44001?

CustomerID	Zip Code	Bought Organic Coffee	Bought Organic Tea
1	44005	Yes	Yes
2	44001	No	No
3	44001	Yes	Yes
4	44005	No	No
5	44003	Yes	No
6	44005	No	Yes
7	44005	No	No
8	44001	No	No
9	44005	Yes	Yes
10	44003	Yes	Yes

○ 0.5

Correct!

◉ 0.3

○ 0.07

○ 0.7

Question 7   1 / 1 pts

Assume that X is a uniformly distributed random variable that takes values from 1 to 40. What is the value of PMF(X=20)?

Correct!

◉ 1/40

○ 20/40

○ 1/20

○ 40/40

Question 8

1 / 1 pts

If $x \sim p_x(x)$ and $y \sim p_y(y)$ are independent, what is $p(x|y) =$?

Correct!

◉ $p_x(x)$

○ 0

○ $p_x(x)p_y(y)$

○ $p_y(y)$

Question 9

1 / 1 pts

Given two sets, A={1, 2, 3, 4, 5}, B={1, 3, 5}. Which of the following is true (choose all that apply)?

☐ A∩B = Φ

☐ A⊂B

Correct!

☑ B⊂A

Correct!

☑ A∩B = B

Question 10

1 / 1 pts

Given matrix $X = \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix}$, what is the X^{-1}?

$\bigcirc$ $\begin{bmatrix} 0 & -0.25 \\ -1 & 0 \end{bmatrix}$

$\bigcirc$ $\begin{bmatrix} -1 & 0 \\ 0 & -4 \end{bmatrix}$

$\bigcirc$ $\begin{bmatrix} 0 & 4 \\ 1 & 0 \end{bmatrix}$

Correct!

$\odot$ $\begin{bmatrix} 1 & 0 \\ 0 & 0.25 \end{bmatrix}$

Question 11

1 / 1 pts

The following figures represent PDFs of normal distributions with different means. Which figure represents the normal distribution with the largest mean?
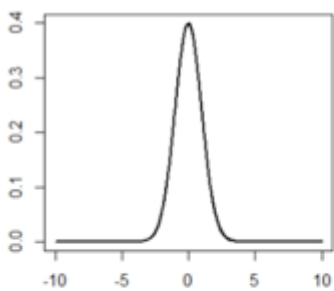
Figure 1
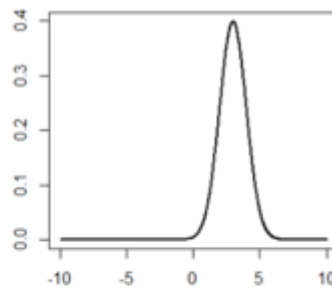
Figure 2
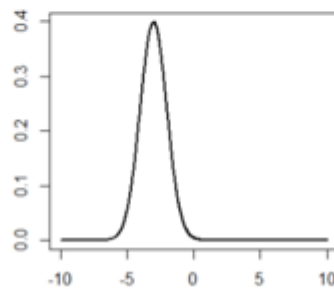
Figure 3

Correct!

$\odot$ Figure 2

○ Figure 3

○ Figure 1

Question 12

1 / 1 pts

In a multivariate Gaussian distribution, if the "Σ" in the PDF is a diagonal matrix, what does it imply?

Correct!

⦿ The features are statistically independent.

○ The features are linearly correlated.

○ The features have the same variance.

○ The features have the same mean.

Quiz Score: **12** out of 12

Mini-Assignment

Due Jan 29 at 11:59pm   **Points** 8   **Questions** 4

Available until Jan 30 at 3am   **Time Limit** None

Allowed Attempts 2

This quiz was locked Jan 30 at 3am.

Attempt History

	Attempt	Time	Score
LATEST	**Attempt 1**	less than 1 minute	8 out of 8

Score for this attempt: **8** out of 8

Submitted Jan 20 at 6:42pm

This attempt took less than 1 minute.

Question 1	2 / 2 pts

Let the entire space be all positive integers smaller than 10. Given A={1, 2, 3, 4, 5}, what is A^c (the complement of A)?

○ {1, 2, 3, 4, 5}

○ Φ (the empty set)

○ {1, 2, 3, 4, 5, 6, 7, 8, 9}

Correct!

◉ {6, 7, 8, 9}

Question 2	2 / 2 pts

The table below shows the purchase history of 10 customers from a set of zip codes that bought organic tea or organic coffee. Using Bayes' Rule, what is the probability that a person who lives in the 44005 zip code and bought organic tea will likely buy the organic coffee?

CustomerID	Zipcode	Bought Organic Coffee	Bought Organic Tea
1	44005	Yes	Yes
2	44001	No	No
3	44001	Yes	Yes
4	44005	No	No
5	44003	Yes	No
6	44005	No	Yes
7	44005	No	No
8	44001	No	No
9	44005	Yes	Yes
10	44003	Yes	Yes

○ 0.5

Correct!

◉ 0.67

○ 0.24

○ 0.3

Question 3   2 / 2 pts

If variable x and z are statistically independent. Which of the following is correct? (select all that apply)

Correct!

☑ Var(x+z) = Var (x)+ Var (z)

☐ Var(x+z) ~= Var (x)+ Var (z)

☐ E(x+z) ~= E(x)+E(z)

Correct!

☑ E(x+z) = E(x)+E(z)

Question 4   **2 / 2 pts**

Consider a wireless cell with four channels . Each channel is in one of two states: busy and available. Both states are equally probable and each channel is independent of any other channel. Define a random variable X to be the number of channels in the busy state. What is the value of E[X]?

Correct!

◉ 2

○ 0.2

○ 1/16

○ 33/16

Quiz Score: **8** out of 8

Mini-Assignment: Supervised learning- Naive Bayes & Logistic Regression

---

Due Feb 7 at 11:59pm **Points** 4 **Questions** 4
Available until Feb 8 at 3am **Time Limit** None
Allowed Attempts 2

Instructions

Unit 3: Supervised learning- Naive Bayes & Logistic Regression

This quiz was locked Feb 8 at 3am.

Attempt History

	Attempt	Time	Score
LATEST	Attempt 1	2 minutes	4 out of 4

Score for this attempt: **4** out of 4
Submitted Feb 5 at 10:26am
This attempt took 2 minutes.

Question 1   **1 / 1 pts**

Let $A_1, A_2, ..., A_n$ be mutually exclusive events that exhaust the probability space Y. Which of the following conditions is true?

○ $\sum_{i=0}^{n} P(Y|A_i) = 1$

Correct! ⦿ $\sum_{i=0}^{n} P(A_i|Y) = 1$

○ $\sum_{i=0}^{n} \sum_{j=0}^{n} P(A_i \cap A_j) = 1$

Question 2

1 / 1 pts

How can logistic regression be used as a classifier?

○ Logistic regression can be used as a classifier by expanding the range of the output of the logistic function.

○ Logistic regression can be used as a classifier by limiting the domain of the logistic function to positive real numbers only.

Correct!

◉ Logistic regression can be used as a classifier by using a threshold on the outcome of the logistic function and using this threshold to classify the inputs.

○ Logistic regression can be used as a classifier by removing the nonlinear relationship between input and output.

Question 3

1 / 1 pts

In Logistic Regression, the parameter η is called the learning rate. What does this parameter control?

○ The size of the final w_0.

○ How quickly new features are incorporated into the model.

○ The speed of convergence of the model.

Correct!

    ◉   The speed at which changes happen to the w parameters.

Question 4   1 / 1 pts

Which of the following plots shows the relationship between the error rates of logistic regression and Naive Bayes when the data size gets larger?

Figure 1

Figure 2

Error Rate (y-axis) vs **Data Size** (x-axis: 100, 1000, 10000)

Legend:
- - - - - Logistic Regression
———— Naïve Bayes
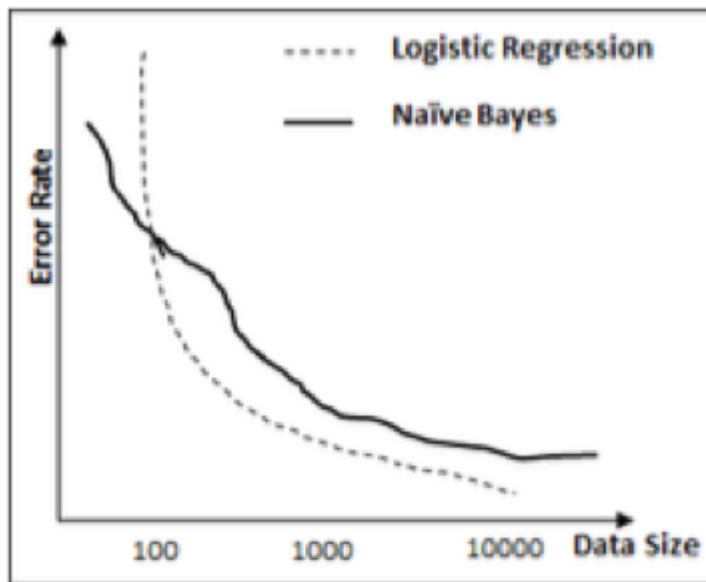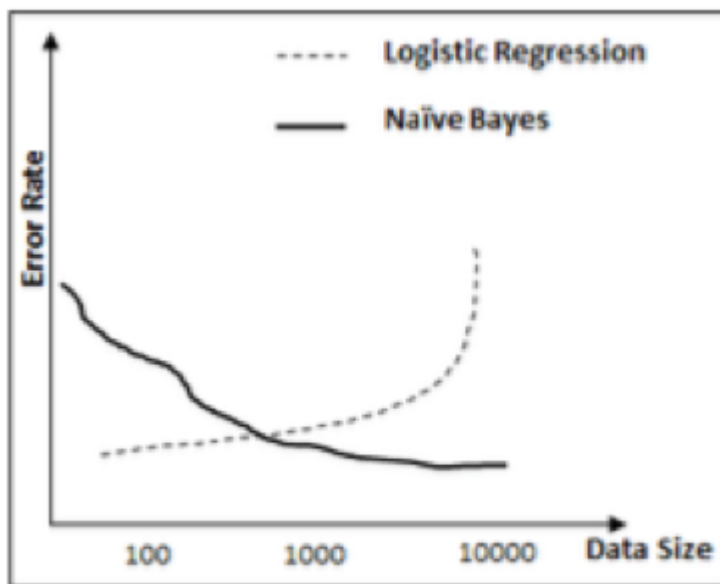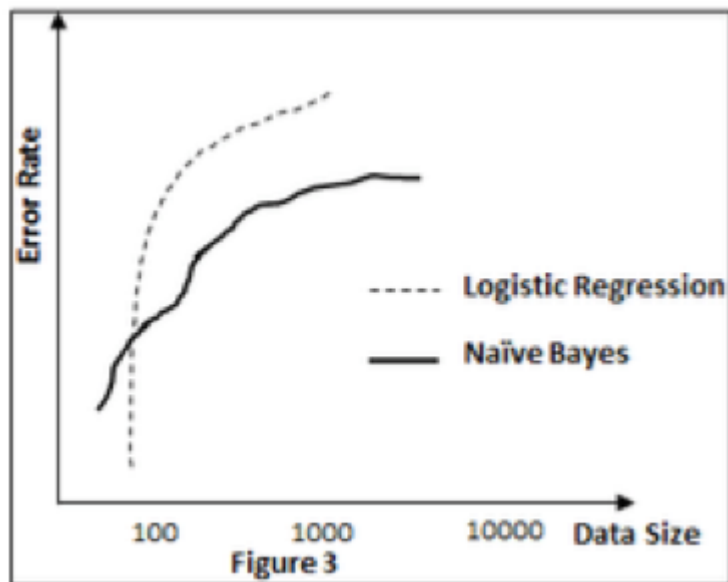
Figure 3

○ Figure 3

○ Figure 2

Correct!

◉ Figure 1

Quiz Score: **4** out of 4

Unit 3: Quiz

Due Feb 10 at 11:59pm   **Points** 7   **Questions** 7

Available until Feb 11 at 3am   **Time Limit** 60 Minutes

This quiz was locked Feb 11 at 3am.

Attempt History

	Attempt	**Time**	**Score**
LATEST	**Attempt 1**	3 minutes	7 out of 7

Score for this quiz: **7** out of 7

Submitted Feb 6 at 2:53pm

This attempt took 3 minutes.

Question 1   1 / 1 pts

Select the answers that best complete the following statement. Unlike Naïve Bayes, which is a ____ model, Logistic Regression is a __model.

Correct!

- ⦿ Generative, discriminative

- ◯ Discriminative, generative

- ◯ Supervised, Unsupervised

- ◯ Quadratic, linear

Question 2   1 / 1 pts

Computing the final values of the parameters for Logistic Regression requires the use of what technique?

Correct!

◉ Gradient Ascent

○ Expectation Maximization

○ Matrix Factorization

Question 3

1 / 1 pts

If the true value of μ is known, then the MLE estimator of

σ^2 is $\sigma^2_{MLE} = \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^2$

Is the estimation of σ^2 unbiased?

Correct!

◉ Yes

○ No

Question 4

1 / 1 pts

Researchers at a medical center are interested in exploring the relationship between patient age (x_1), patient weight (x_2) and the presence (1) or absence (0) of a particular disease. If researchers decide to use logistic regression, which of the following would be interpreted as the probability that the positive outcome (e.g., disease) is present?

○ $\text{Probability} = \beta_0 + \beta_1 x_1 + \beta_2 x_2$

    ○   Probability $= \frac{e^{\beta_0+\beta_1 x_1+\beta_2 x_2}}{1-e^{\beta_0+\beta_1 x_1+\beta_2 x_2}}$

Correct!   ◉   **Probability** $= \frac{e^{\beta_0+\beta_1 x_1+\beta_2 x_2}}{1+e^{\beta_0+\beta_1 x_1+\beta_2 x_2}}$

    ○   Probability $= \frac{e^{-\beta_0-\beta_1 x_1-\beta_2 x_2}}{1+e^{-\beta_0-\beta_1 x_1-\beta_2 x_2}}$

Question 5   1 / 1 pts

Suppose the scores of randomly selected students are normally distributed with an unknown mean and a standard deviation. A random sample of 10 students returns the following scores: [81, 71, 71, 74, 56, 92, 83, 74, 91, 66]. Estimate μ by using maximum likelihood estimation.

(Choose the closest one.)

○ 56

○ 92

○ 11

Correct!   ◉ 76

Question 6   1 / 1 pts

When tossing a dice several times, let α_0 stand for the number of odd rolls and α_e stand for the number of even rolls. Let p be the probability of getting an even number. Using maximum likelihood estimation, how do you estimate p?

Correct!   ◉ $\hat{p} = \frac{\alpha_e}{\alpha_e + \alpha_0}$

$\hat{p} = \frac{\alpha_e}{\alpha_0}$

$\hat{p} = \frac{\alpha_0}{\alpha_e + \alpha_0}$

$\hat{p} = \frac{\alpha_0}{\alpha_e}$

Question 7   1 / 1 pts

Which of the following is true for generative or discriminative classifier models that read the input x and the label y? (Select all that apply.)

Correct!

☑ A generative classifier makes its predictions by using Bayes' rule to calculate P(y|x).

☐ A discriminative classifier learns a model of the joint probability p(x, y).

Correct!

☑ A discriminative classifier models the posterior P(y|x) directly.

Correct!

☑ A generative classifier attempts to learn p(x|y).

☐ A discriminative classifier learns a model of the joint density p(x, y).

☐ No answer text provided.

Quiz Score: **7** out of 7

Mini-Assignment: SVMs

Due Feb 26 at 11:59pm   **Points** 8   **Questions** 4
Available until Feb 27 at 2:59am   **Time Limit** 60 Minutes
Allowed Attempts 2

This quiz was locked Feb 27 at 2:59am.

Attempt History

	Attempt	Time	Score
LATEST	**Attempt 1**	1 minute	8 out of 8

Score for this attempt: **8** out of 8
Submitted Feb 14 at 7:21pm
This attempt took 1 minute.

Question 1	2 / 2 pts

Given the following dataset in the 1-d space, which consists of 3 positive data points{-1,0,1} and 3 negative data points{3,-2,-2}.
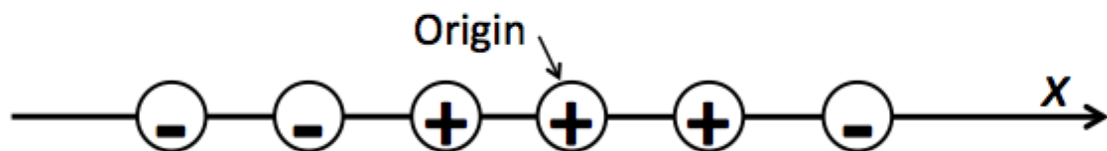
Are they linearly separable in original 1-d space?

○ Cannot tell

Correct!   ◉ No

○ Yes

Question 2

2 / 2 pts

Given a hyperplane defined by $w^t x + b = 0,$ what is the distance of the origin to this hyperplane?

○ 1/b

○ b

Correct!

◉ abs(|b/||w||)

○ abs(1/b)

Question 3

2 / 2 pts

The following table contains seven observations in two dimensions, $X1$ and $X2$. Each observation has an associated class label, Y: Green and Yellow. The observations are plotted in a 2-Dimensional space. What is the equation for the maximal margin separating hyperplane?

X_1	X_2	Y
3	4	Green
2	2	Green
4	4	Green
1	4	Green
2	1	Yellow
4	3	Yellow
4	1	Yellow
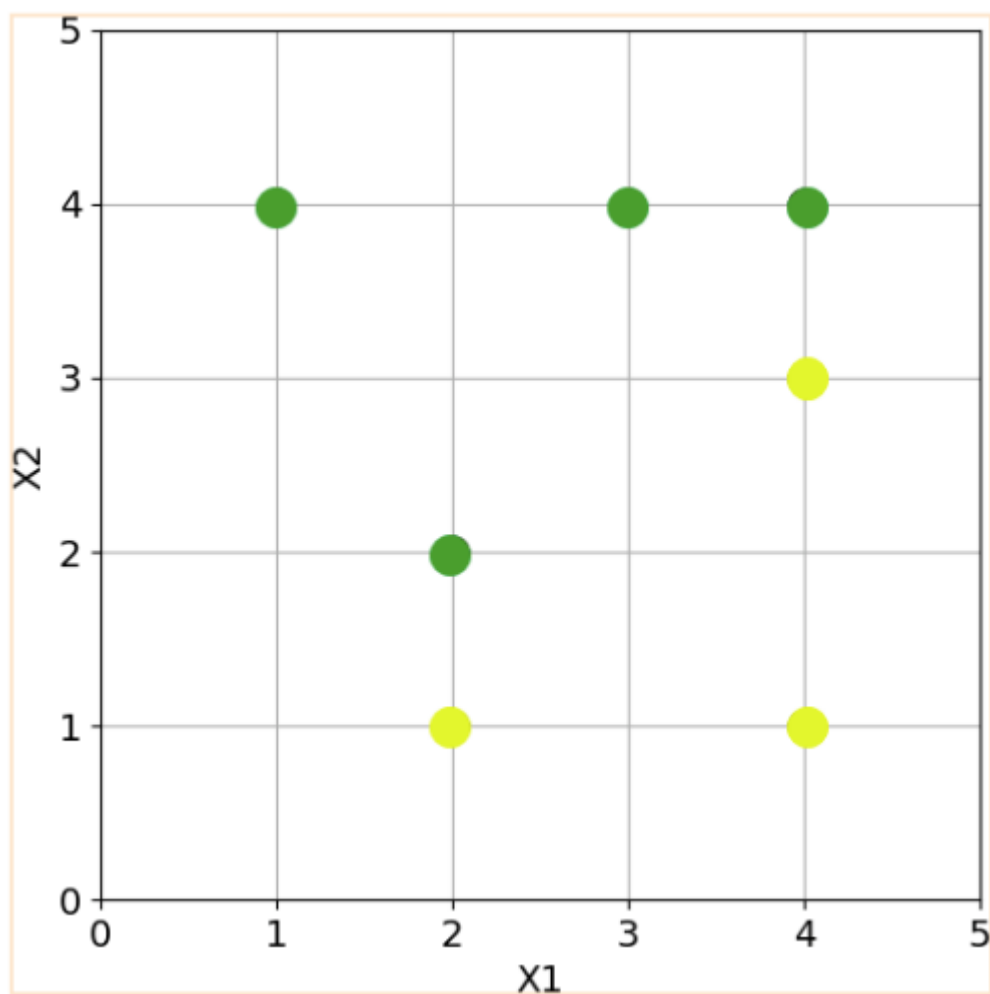
○ ½ + X1+X2= 0

○ 2 - X1+X2= 0

Correct!

◉ ½ - X1+X2= 0

○ ½ + X1-X2= 0

Question 4

2 / 2 pts

Given the following feature space with two classes (plus and minus), which hyperplane represents the best choice for a hyperplane?
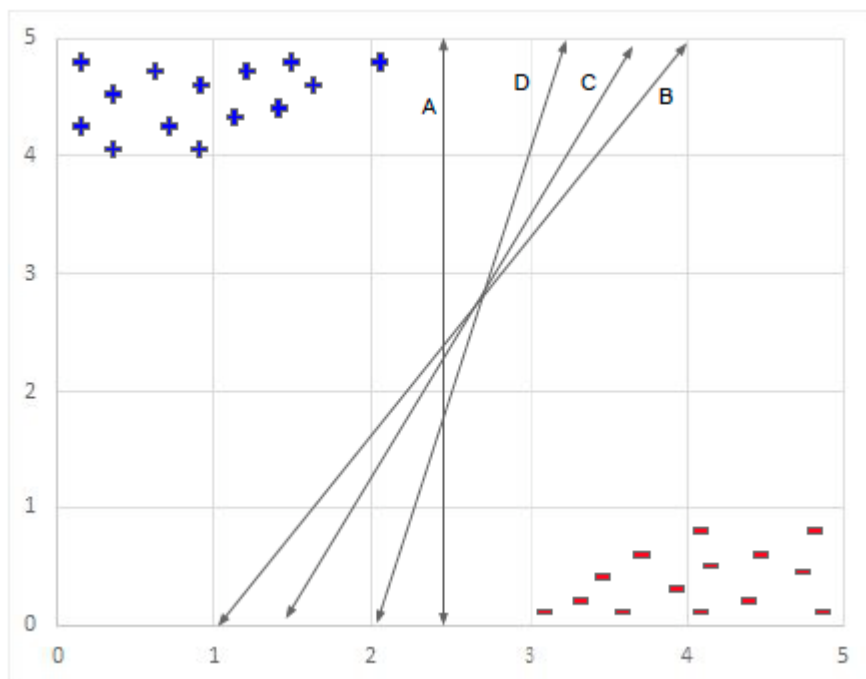
Correct!

◉ y = 5/3*x - 5/3

○ y = 2x -4

○ y = 4/3*x - 4/3

○ y = x-1

Quiz Score: **8** out of 8

Unit 4: Quiz

Due Feb 26 at 11:59pm   **Points** 9   **Questions** 9
Available until Feb 27 at 2:59am   **Time Limit** 60 Minutes

This quiz was locked Feb 27 at 2:59am.

Attempt History

	Attempt	Time	Score
LATEST	Attempt 1	6 minutes	9 out of 9

Score for this quiz: **9** out of 9
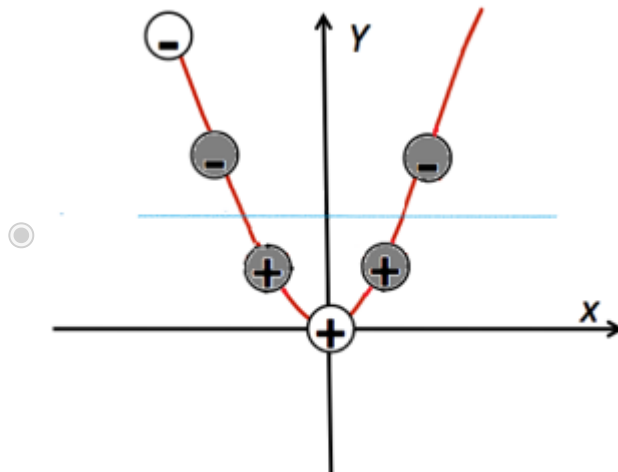Submitted Feb 14 at 7:27pm
This attempt took 6 minutes.

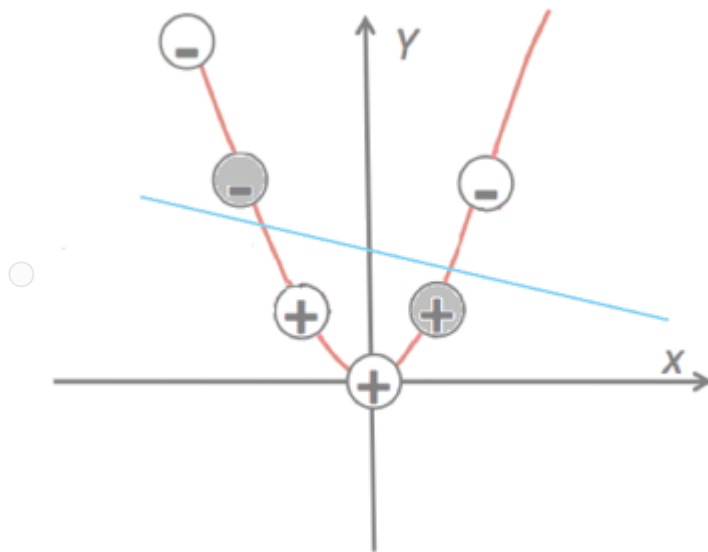| Question 1	1 / 1 pts

Suppose we find a feature mapping from the 1-d space to the 2-d space, and the mapped points are shown in the following figure. If the decision boundary is given by hard-margin linear SVM, which figure shows the correct decision boundary and support vector(s)?
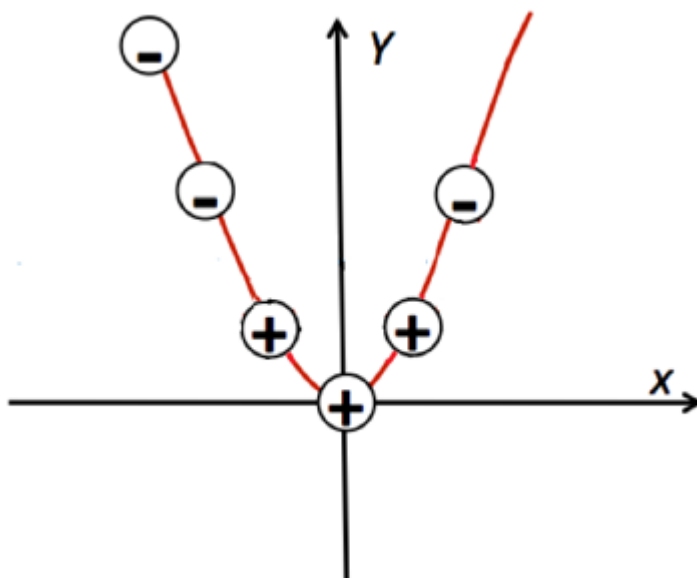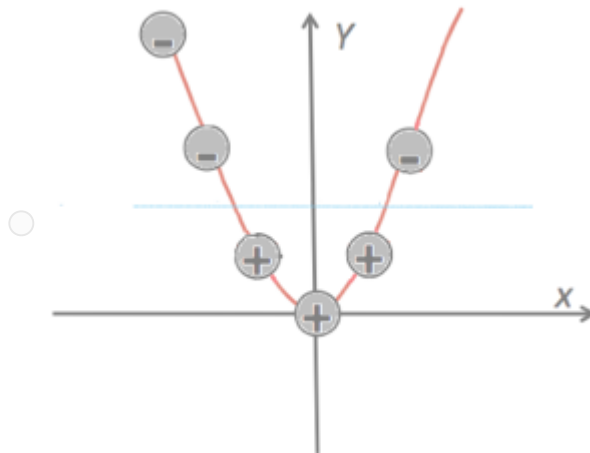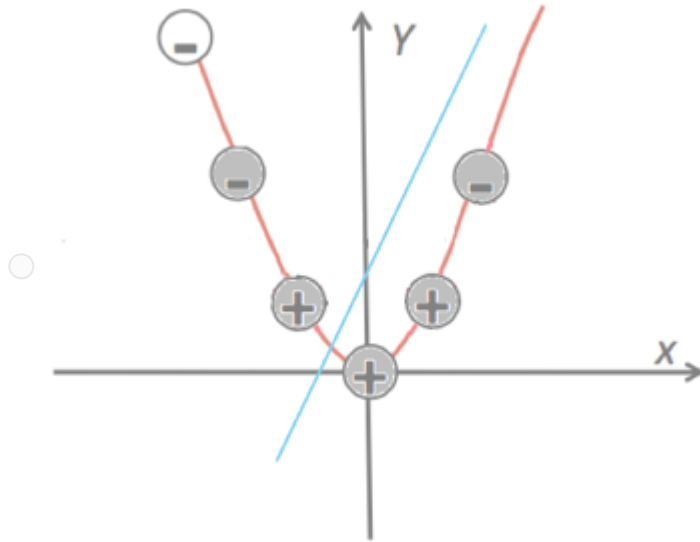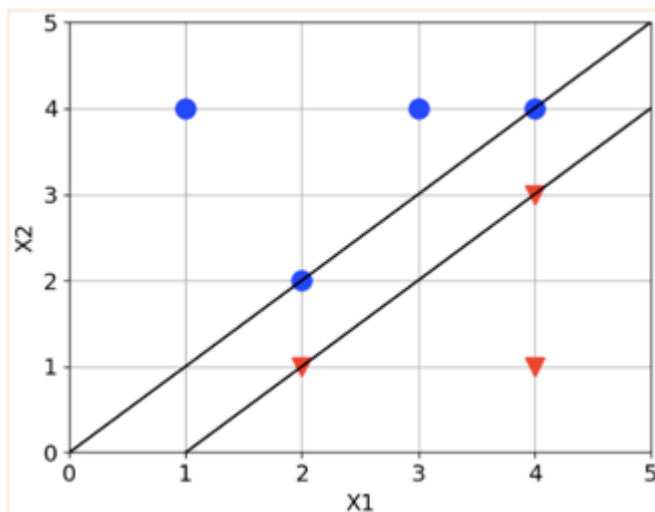
Correct!

The decision boundary in this example should classify all the samples correctly, and the support vectors are the samples that are closest to the decision boundary.

Question 2

1 / 1 pts

A maximal margin classifier found the following canonical hyperplanes. What is the value of margin?

○ $\sqrt{2}$

Correct!

⦿ $\dfrac{\sqrt{2}}{2}$

◯ ¼

◯ ½

> The margin is the shortest distance between the two lines passing through the support vectors.

Question 3

1 / 1 pts

Consider the figure above. What is the classification rule for the maximal margin classifier that classifies the points in blue (the circles)
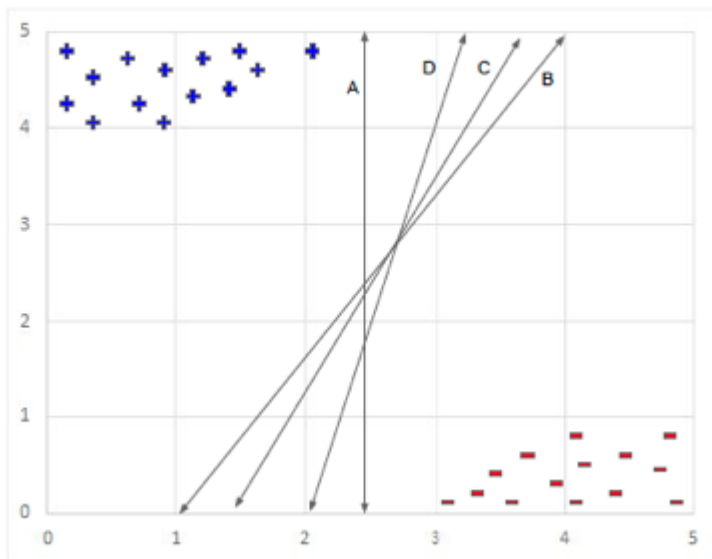
Correct!

⦿ $1 - 2x_1 + 2x_2 > 0$

◯ $1 - x_1 + x_2 > 0$

◯ $1 - x_1 + x_2 < 0$

◯ $1 + x_1 - 2x_2 > 0$

> The classifier will be at the middle of the two lines passing through the support vectors.

Question 4

1 / 1 pts

Given this feature space with two classes (plus and minus), what is the formula for the best choice of hyperplane, in y = mx+b format?

○ C

○ D

Correct!

◉ B

○ A
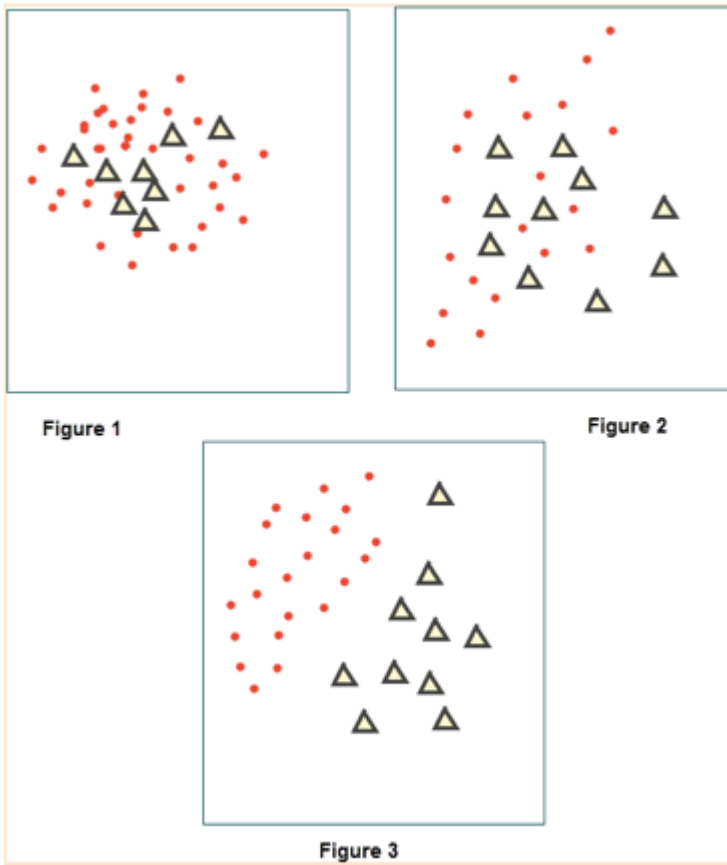

Question 5   1 / 1 pts

Which of the following data sets do NOT require a kernel transformation to transform data into higher dimensions where it can be separated with a hyperplane?

Figure 1   Figure 2

Figure 3

○ Figure 1

Correct!   ◉ Figure 3

○ Figure 2

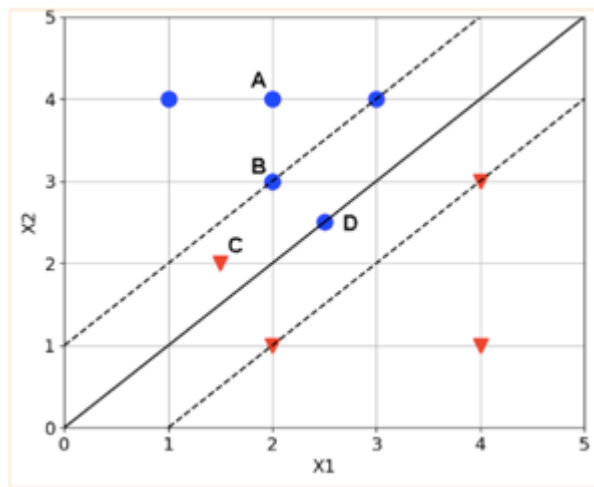If the samples are not linearly separable, you can use the kernel trick.

Question 6

1 / 1 pts

Which of the following is a point on the hyperplane $1 + 2X_1 + 2X_2 = 0$?

○ (1/2,-1/4)

○ (0,-1)

Correct!   ◉ (-1,1/2)

○ (-1/2,-1/2)

Question 7   **1 / 1 pts**

In the following application, the SVM classifier is applied to a non-linearly separable data. For which points is ξ=0? (Select all that apply.)
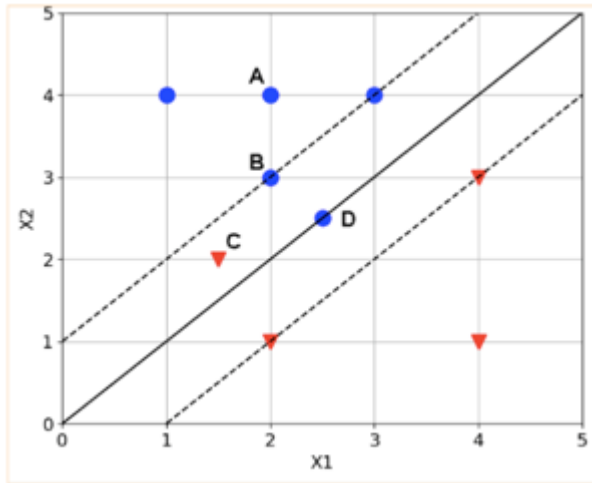

☐ Point D

☐ Point C

Correct!   ☑ Point A

Correct!   ☑ Point B

The slack is not zero if the sample is between the two dotted line.

Question 8

1 / 1 pts

In the following feature space with two classes (in blue and red respectively),

the SVM classifier is applied to a non-linearly separable data. For which points is $\xi > 1$?

○ Point B

○ Point A

Correct!

◉ Point C

○ Point D

The sample which is misclassified, will have a slack > 1

Question 9

1 / 1 pts

Which of the following is true for SVMs? (Select all that apply.)

☐ For linearly separable data of two classes, there's only one hyperplane that can correctly classify all samples.

Correct!   ☑ The support vectors alone define the decision boundary in SVM.

Correct!   ☑
SVMs use the kernel trick to effectively transform the problem into a space where the data is linearly separable.

Quiz Score: **9** out of 9

Mini-Assignment: Bayesian Network & HMM

Due Mar 4 at 11:59pm   **Points** 5   **Questions** 5
Available until Mar 5 at 2:59am   **Time Limit** None
Allowed Attempts 2

Instructions

Unit 5: Bayesian Network & HMM

This quiz was locked Mar 5 at 2:59am.

Attempt History

	Attempt	Time	Score
LATEST	**Attempt 1**	2 minutes	5 out of 5

Score for this attempt: **5** out of 5
Submitted Mar 2 at 2:06pm
This attempt took 2 minutes.

Question 1	**1 / 1 pts**

In HMM, to find out what is the most probable path of the states, given an observation sequence O, which algorithm should be used?

    ◯   Baum-Welch

    ◯   EM

    ◯   Forward-Backward

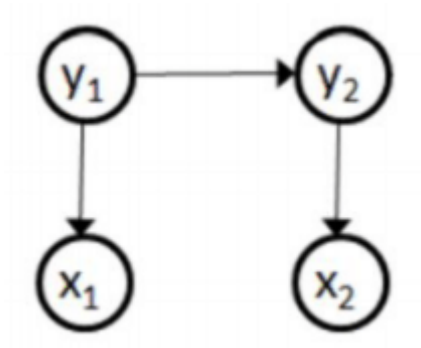**Correct!**   ◉   Viterbi Decoding

Question 2   1 / 1 pts

Given an HMM Λ={Θ,Ω,A,B,π}, which matrix represents the initial state?

○ A

○ Θ

○ B

Correct!   ◉ π

Question 3   1 / 1 pts

Consider the following hidden Markov model which consists of two observations (x1, x2) and two hidden states (y1, y2). What is the joint probability distribution P(x1, x2, y1, y2)?

○ P(y1)P(x1|y1) P(x2|y2)

○ P(x1|y1) P(x2|y2)

○ P(x1|y1)P(y2|y1)P(x2|y2)

Correct!   ◉ P(y1)P(x1|y1)P(y2|y1)P(x2|y2)

Question 4

1 / 1 pts

Given the following HMM model, where symbols X, Y and Z represent the possible observations in the states. In a circle representing a state, the symbols are equally possible observations in that state. If the system can only start equally likely from state 1 or state 2, what is the π?

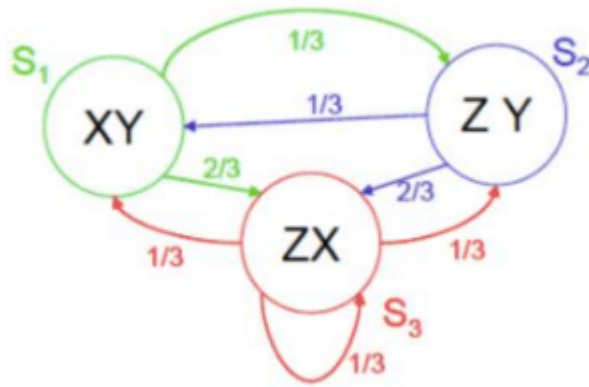- ○ π 1=1/3, π 2=1/3, π 3=1/3

Correct!

- ◉ π 1=1/2, π 2=1/2, π 3=0

- ○ π 1=1, π 2=0, π 3=0

- ○ π 1=0, π 2=1, π 3=0
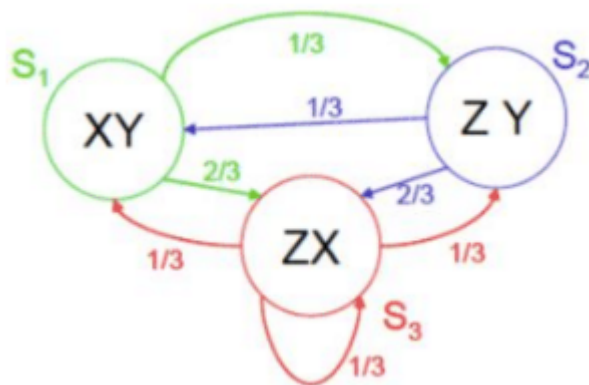
Question 5

1 / 1 pts

Given the following HMM model, where symbols X, Y and Z represent the possible observations in the states. In a circle representing a state, the symbols are equally possible observations in that state. What is the value of a_{33} in the state transition probability matrix?

○ 2/3

Correct!   ◉ 1/3

○ 1

○ 0

Quiz Score: **5** out of 5

Unit 5: Quiz

Due Mar 4 at 11:59pm   **Points** 8   **Questions** 8
Available until Mar 5 at 2:59am   **Time Limit** 60 Minutes

This quiz was locked Mar 5 at 2:59am.

Attempt History

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

Score for this quiz: **8** out of 8
Submitted Mar 2 at 2:20pm
This attempt took 10 minutes.

Question 1   1 / 1 pts

Given the following HMM model, where symbols X, Y and Z represent the possible observations in the states. In a circle representing a state, the symbols are equally possible observations in that state. What is the value of $b_3(Z)$ in the observation probability matrix?

- ○ 0

- ○ 1/3

Correct!
- ◉ 1/2
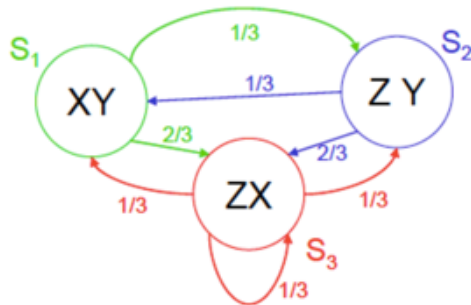
- ○ 2/3
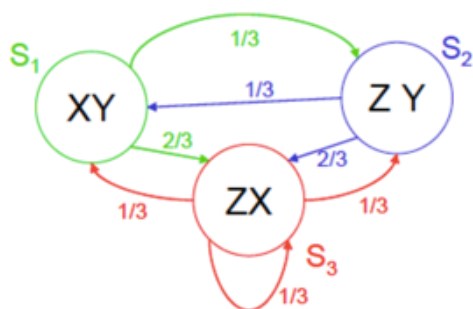
Question 2

1 / 1 pts

Given the following HMM model, where symbols X, Y and Z represent the possible observations in the states. In a circle representing a state, the symbols are equally possible observations in that state. Suppose π 1 =1/2, what is the value of P(Q), Q = S1S3S3?

○ 1/2

○ 2/3

Correct! ◉ 1/9

○ 1/3

Question 3

1 / 1 pts

Given the following HMM model, where symbols X, Y and Z represent the possible observations in the states. In a circle representing a state, the symbols are equally possible observations in that state. How much is P(O1=Y, O2=X, O3=Z| S2S3S3)?
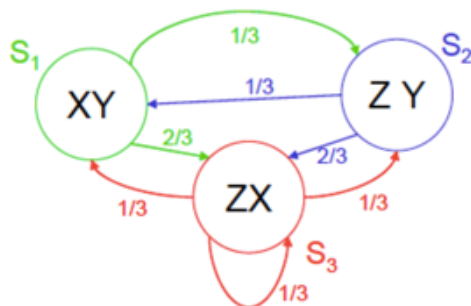

Correct! ◉ 1/8

○ 2/3

○ 1/2

○ 1/4

Question 4

1 / 1 pts

Given the following HMM model, where symbols X, Y and Z represent the possible observations in the states. In a circle representing a state, the symbols are equally possible observations in that state. How many states are there in the model?

○ 2

Correct!

◉ 3

○ 1

○ 4
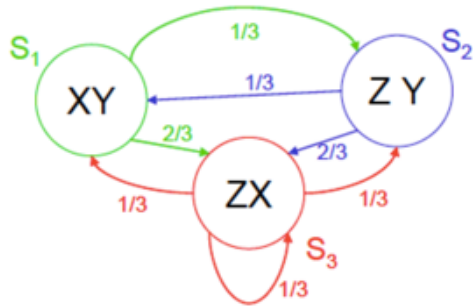
Question 5

1 / 1 pts

The Viterbi Algorithm for state prediction is a dynamic programming solution.

Correct!

◉ True

○ False

Question 6   **1 / 1 pts**

Given an HMM $\Lambda=\{\Theta,\Omega,A,B,\pi\}$, which matrix represents the set of output (observation)?

Correct!

○ A

◉ Ω

○ Θ

○ B

Question 7   **1 / 1 pts**

To learn the probabilities in graphical models, which of the following may be used? (Select all that apply

Correct!

☑ Use the MLE principle for estimation.

Correct!

☑ Make proper assumption about the priors.

Correct!

☑ Use relative frequency to estimate probability.

---

Question 8   **1 / 1 pts**

Given the following HMM model, where symbols X, Y and Z represent the possible observations in the states. In a circle representing a state, the symbols are equally possible observations in that state.
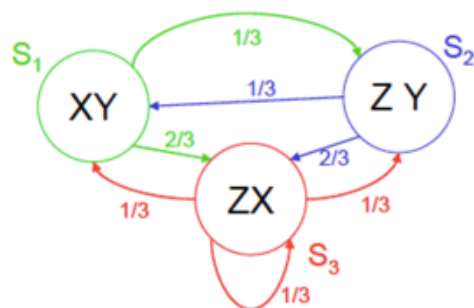
Suppose that the first state is S1, and we have a path of length 100, how many different possible state paths are there in total?

○ 3^100

Correct!
   ⦿ 3^99

○ 4^99

○ 4^100

Quiz Score: **8** out of 8

Unit 6: Quiz

Due Mar 20 at 11:59pm   **Points** 10   **Questions** 10

Available until Mar 21 at 2:59am   **Time Limit** 60 Minutes

This quiz was locked Mar 21 at 2:59am.

Attempt History

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

Score for this quiz: **10** out of 10

Submitted Mar 20 at 10:07pm

This attempt took 7 minutes.

Question 1   1 / 1 pts

K-Means algorithm is not for solving _____ problems. (Select all that apply.)

☐ Clustering

Correct!   ☑ Logistic Regression

Correct!   ☑ Linear Regression

Correct!   ☑ Classification

Question 2   1 / 1 pts

Which of the following defines hard clustering?

Correct!

○ Every data point falls into one cluster and one only.

○ Every cluster contains one data point only.

○ The entire dataset is clustered in one cluster only.

○ The cluster's centers never change.

Question 3   1 / 1 pts

Which of the following can NOT be used as a stopping criterion for the K-means algorithm? (select all that apply)

Correct!

☑ The sum of squared error reaches a local maximum.

Correct!

☑ There are points that belong to more than one cluster.

☐ The centroids have stopped changing.

Correct!

☑ Some centroids contain no point at all after an iteration.

> "Some centroids contain no point at all after an iteration." is not a stopping criterion. You should randomly assign a point (which is not a centroid) back to that cluster. Another possible solution is that, you keep that cluster empty and return less centroids. In both cases you don't stop the iteration.

Question 4   1 / 1 pts

Which of the following reasons is **not** the reason why the K-means algorithm will likely end up with sub-optimal clustering? (Select all that apply.)

   ☐ Bad choices for the initial cluster centers.

Correct!

   ☑

Choosing a k that corresponds to the number of natural clusters in the dataset.

Correct!

   ☑ Fast convergence of the K-means algorithm.

Correct!

   ☑ Existence of closely-located data samples in the dataset.

Question 5

1 / 1 pts

Which of the following is a step in K-means algorithm implementation? (Select all that apply.)

Correct!

   ☑ Set initial cluster centers.

Correct!

   ☑

Compute the distance between every data point and each cluster center.

Correct!

   ☑ Assign each data point to its nearest cluster center.

   ☐ Initialize the cluster centers to random points at each iteration.

Question 6

1 / 1 pts

Which of the following is true for the K-means algorithm? (Select all that apply.)

Correct!

☑ The value of K needs to be set before running the algorithm.

Correct!

☑

On every iteration of the algorithm, the cost function should either stay the same or decrease (in particular, it should not increase).

☐

Once a data point has been assigned to a particular center, it will no longer be re-assigned to a different center in later iterations.

☐

The algorithm will produce the same set of centers regardless of the initialization.

Question 7

1 / 1 pts

Given a data point, x_1 = (1,1), and two cluster centers, C_1 = (1,0) and C_2 = (0,0), which cluster does x_1 belong to?

○ Cluster 2

Correct!

⦿ Cluster 1

○ Both clusters 1 and 2

○ Neither

Question 8

1 / 1 pts

You are given 5 data points in a 1-D space:
$x_1 = -5, x_2 = -4, x_3 = 0, x_4 = 2$ and $x_5 = 3$. Suppose K = 2 and
the initial cluster centers are $\mu_1 = 0$ and $\mu_2 = 1$. When running K-means, what are the final cluster centers after convergence?

- ○ $\mu_1 = 0, \mu_2 = 1$

- ○ $\mu_1 = -3, \mu_2 = 2.5$

- ○ $\mu_1 = -4.5, \mu_2 = 2.5$

Correct!

- ⦿ $\mu_1 = -4.5, \mu_2 = 5/3$

Question 9   1 / 1 pts

K-means clustering is applied with K = 2, on a small dataset of six points
as illustrated below. What are the centroids for each cluster?
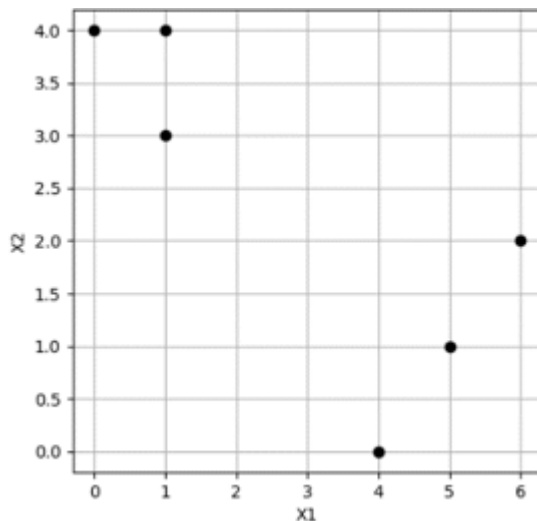
Correct!

- ⦿ (2/3,11/3) and (5,1)

- ○ (1,4) and (5,3)

- ○ (11/3, 2/3) and (5,1)

    ○ (4,0) and (2,3)

Question 10   1 / 1 pts

The following points (1,1), (2,6), (3,3),(3,4),(4,2),(6,4),(5.3) are clustered using the K-means algorithm.

The assignment step of the K-Means algorithm creates the following clusters: cluster_1: {(1,1),(2,6),(3,4)} and cluster_2: {(3,3),(4,2),(6,4), (5,3)}.

What are the new centroids of each cluster?

Correct!

    ◉ (2,11/3),(4.5,3)

    ○ (2,6),(5,3)

    ○ (11/3,2),(3,4.5)

    ○ (11/3,4.5),(2,3)

Quiz Score: **10** out of 10

Mini-Assignment: K-Means

Due Mar 20 at 11:59pm   **Points** 18   **Questions** 9

Available until Mar 21 at 2:59am   **Time Limit** None

Allowed Attempts 2

This quiz was locked Mar 21 at 2:59am.

Attempt History

	Attempt	Time	Score
KEPT	**Attempt 2**	1 minute	17 out of 18
LATEST	**Attempt 2**	1 minute	17 out of 18
	Attempt 1	2 minutes	16.67 out of 18

Score for this attempt: **17** out of 18
Submitted Mar 20 at 10:18pm
This attempt took 1 minute.

Question 1	**2 / 2 pts**

Which of the following should be taken into consideration in order to obtain good results when using the K-means algorithm? (select all that apply)

☐ Standard deviation of input features at each iteration.

Correct!   ☑ Selection of the initial cluster centers

Correct!   ☑ Selection of the number of clusters

Correct!   ☑ Existence of the outliers

Question 2   2 / 2 pts

Which of the following is true for K-Means clustering? (Select all that apply)

Correct!
☑ The overall goal is to minimize the total squared distance from all points of their cluster centers.

☐ The final clusters are not sensitive to the initial cluster centers.

Correct!
☑ There is no guarantee that it reaches the global optimum.

Correct!
☑ Completely different clusters may arise from small changes in the initial random choice.

Question 3   1 / 2 pts

Which of the following is one of the properties of the K-means algorithm? (Select all that apply)?

Correct!
☑ It finds a local optimum.

☐ No matter which initial cluster centers are selected, it always ends up with same clusters.
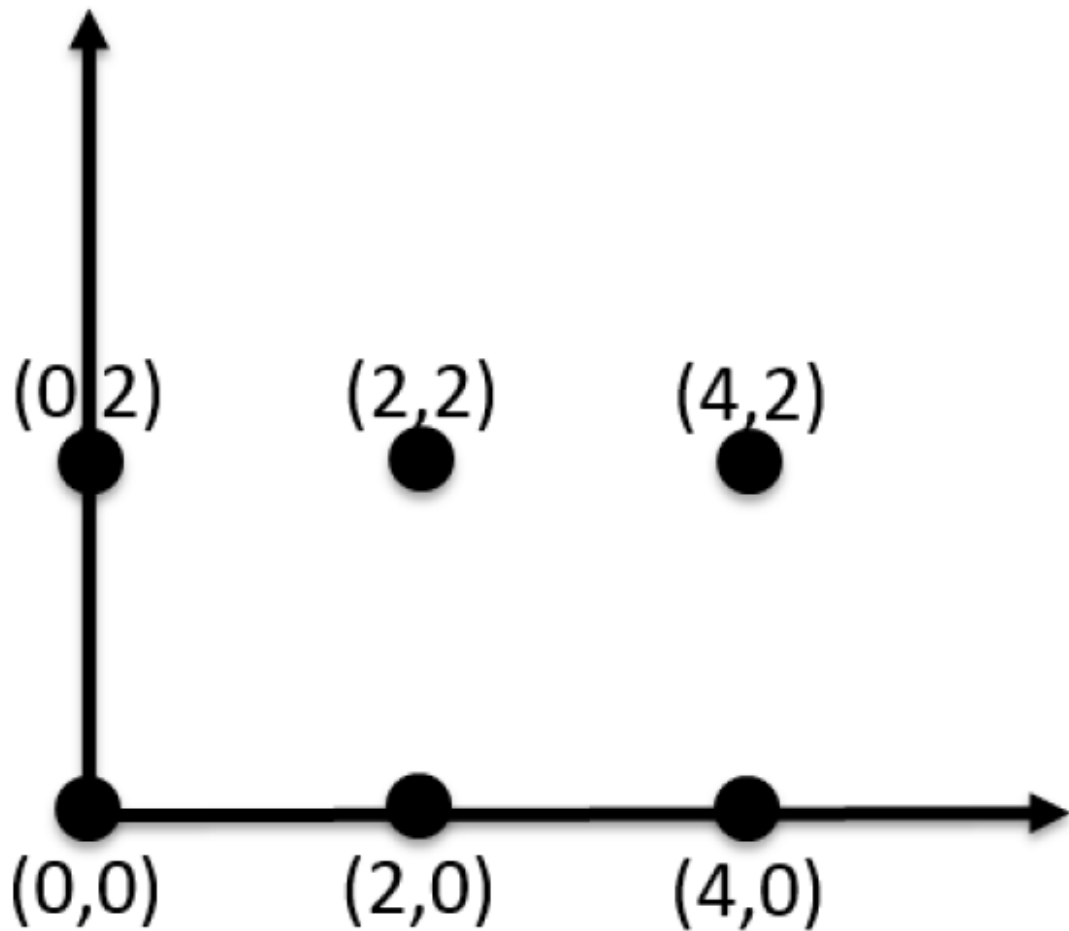
orrect Answer
☐ The initial choice of cluster centers may have a large effect on the result.

☐ It converges for most cases but not always.

Question 4	2 / 2 pts

Suppose we run K-means on the following dataset with six data points to find two clusters.

(0,2)   (2,2)   (4,2)

(0,0)   (2,0)   (4,0)

Suppose the initial cluster centers are (0,0) and (5,0). How many iterations does the algorithm take until convergence?

○ The algorithm cannot converge.

○ It is not possible to determine.

○ At least 5

Correct!

○ Less than 3

Question 5

2 / 2 pts

Following the previous question. What is the cluster assignment for each data point after iteration 1?

○ bottom three belong to the first cluster, and the upper three belong to the second cluster.

○ No answer text provided.

○ Left two belong to the first cluster, and the right four belong to the second cluster.

Correct!

◉ Left four belong to the first cluster, and the right two belong to the second cluster.

Question 6

2 / 2 pts

Following the previous question. What are new cluster centers after one iteration of update?

Correct!

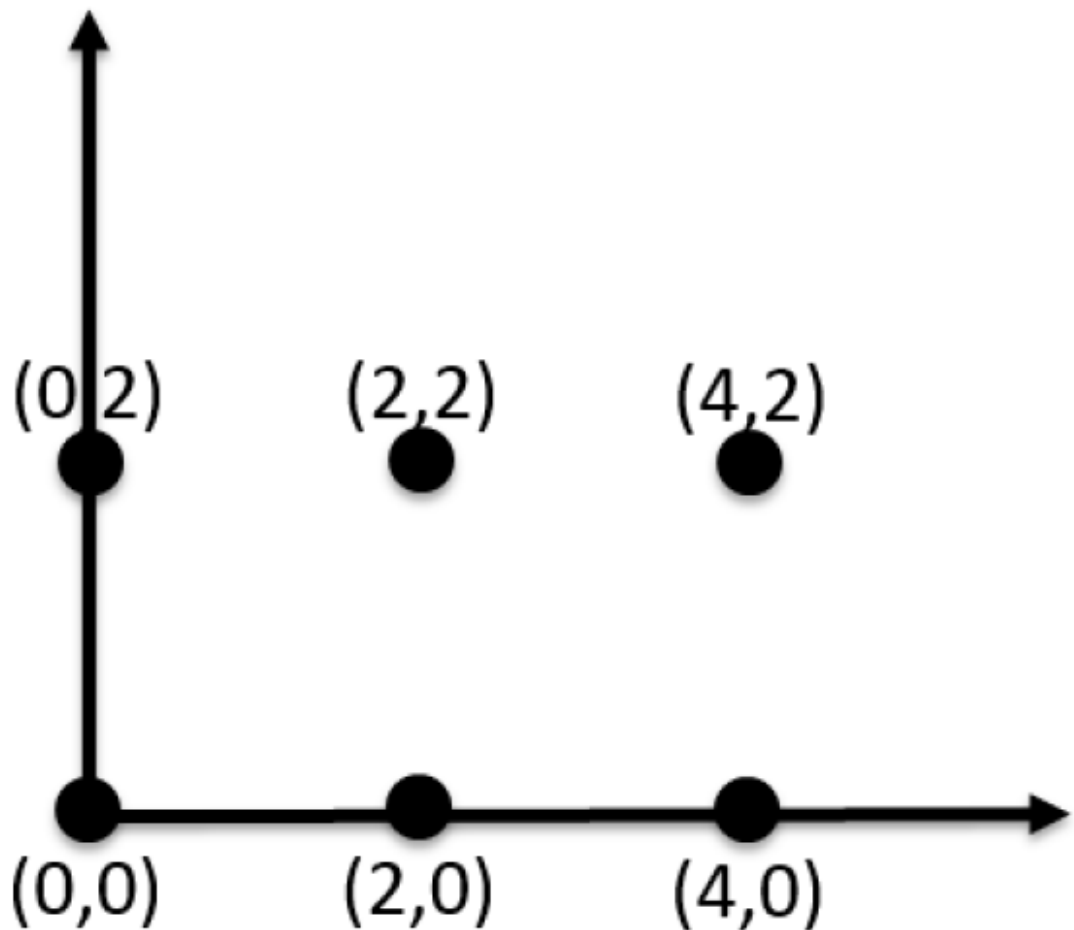◉ (1,1) and (4,1)

○ (1,1) and (3,1)

○ (1,1) and (5,1)

○ (0,0) and (3,1)

Question 7

2 / 2 pts

We run k-means on the following dataset with six data points to find two clusters. If the initial cluster centers are (2,0) and (2,2), How many iterations does the algorithm take until convergence?

Correct!

◉ Less than 3

○ It is not possible to determine.

○ More than 3

○ The algorithm cannot converge.

Question 8   2 / 2 pts

Following the previous question. What is the cluster assignment for each
data point after Step1?

○ No answer text provided.

○ Left two belong to the first cluster, and the right four belong to the second
cluster.

Correct!

◉ bottom three belong to the first cluster, and the upper three belong to the
second cluster.

○ Left four belong to the first cluster, and the right two belong to the second
cluster

Question 9   2 / 2 pts

Following the previous question. What are new cluster centers after one
iteration?

Correct!

◉ (2,0) and (2,2)

○ (1,1) and (2,1)

○ (2,2) and (3,1)

○ (2,0) and (4,1)

Quiz Score: **17** out of 18

Unit 7: Quiz

Due Mar 30 at 11:59pm   **Points** 5   **Questions** 5

Available until Mar 31 at 2:59am   **Time Limit** 60 Minutes

This quiz was locked Mar 31 at 2:59am.

Attempt History

	Attempt	Time	Score
LATEST	**Attempt 1**	2 minutes	5 out of 5

Score for this quiz: **5** out of 5

Submitted Mar 29 at 5:56pm

This attempt took 2 minutes.

Question 1   1 / 1 pts

Which algorithm would we use to get the decision boundary shown in the following figure?

    ○ MinMaxCut

Correct!   ● **MinCut**

    ○ RatioCut

    ○ Ncut

Question 2

1 / 1 pts

Which of the following approach considers both inter-cluster and intra-cluster similarity? (select all that apply)

☐ MinCut

Correct! ☑ MinMaxCut

☐ RatioCut

Correct! ☑ Ncut

Ncut also consider intra similarity because when the objective function encourages Vol(A) and Vol(B) to be large. In such case it encourages the bonding within each cluster to be strong.

Question 3

1 / 1 pts

Using 2-way partitioning recursively to get k-way partition is inefficient and may cause stability issues.

Correct! ◉ True

○ False

Question 4

1 / 1 pts

Given a graph with 6 nodes (i.e., data points) in the following figure, we want to run the spectral clustering for MinCut to find two clusters.
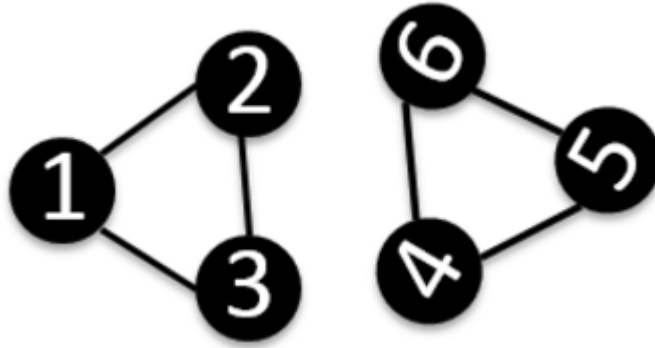
Which of following is the *Laplacian* matrix?

[2,0,0,0,0,0;

0,2,0,0,0,0;

0,0,2,0,0,0;

0,0,0,2,0,0;

0,0,0,0,2,0;

○ 0,0,0,0,0,2];

Correct!

[2,-1,-1,0,0,0;

-1,2,-1,0,0,0;

-1,-1,2,0,0,0;

0,0,0,2,-1,-1;

0,0,0,-1,2,-1;

◉ 0,0,0,-1,-1,2];

[2,1,1,0,0,0;

1,2,1,0,0,0;

1,1,2,0,0,0;

0,0,0,2,1,1;
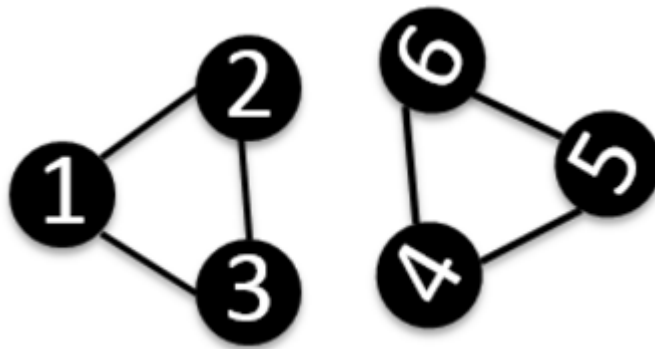
0,0,0,1,2,1;

○ 0,0,0,1,1,2];

[0,1,1,0,0,0;

1,0,1,0,0,0;

1,1,0,0,0,0;

0,0,0,0,1,1;

0,0,0,1,0,1;

○   0,0,0,1,1,0];

Question 5   1 / 1 pts

Given a graph with 6 nodes (i.e., data points) in the following figure, we want to run the spectral clustering for MinCut to find two clusters.

What is the cut size?

○   1

○   3

Correct!   ◉   0

○   2

In an unweighted undirected **graph**, the **size** or weight of a **cut** is the number of edges crossing the **cut**. In a weighted **graph**, the value or weight is defined by the sum of the weights of the edges crossing the **cut**

Quiz Score: **5** out of 5

Mini-Assignment: Dimensionality Reduction

Due Apr 7 at 11:59pm   **Points** 8   **Questions** 4
Available until Apr 8 at 2:59am   **Time Limit** 30 Minutes
Allowed Attempts 2

This quiz was locked Apr 8 at 2:59am.

Attempt History

	Attempt	Time	Score
LATEST	**Attempt 1**	less than 1 minute	8 out of 8

Score for this attempt: **8** out of 8
Submitted Apr 7 at 6:48pm
This attempt took less than 1 minute.

Question 1   **2 / 2 pts**

Suppose that we are given a data set of 20 points (with zero mean) in 2d space. When the first principal component of the data set is computed, the result is as follows:

$$u = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right)$$

If we add one data point to the data set at (2,2), how would this first principle component change?

○ It would rotate toward the y axis.

Correct!   ● It would not.

○ It would rotate toward the x axis.

○ It would extend in the same direction.

Question 2

2 / 2 pts

Suppose that we are given a data set of 20 points (with zero mean and the norm of each point is smaller than 1) in 2d space. When the first principal component of the data set is computed, the result is as follows:

$$u = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right)$$

If we add an extremely large number of data points to the data set at (0, -5), how would this change the principle component?

○ It would towards (1, 0)

○ It would towards (-1, 0)

Correct!

◉ It would towards (0, -1)

The direction will towards (0,-5)

Question 3

2 / 2 pts

Which of the following is one way to find out whether the dimensionality reduction algorithm performs well?

○ The dimension of the data matrix

○ The number of principal components.

Correct!

⦿ The measure of the reconstruction error.

◯ The length of the principal component vectors

Question 4   2 / 2 pts

The PCA makes which of the following assumptions about the data?

◯ The data has a zero standard deviation.

Correct!

⦿ The data has a zero mean

◯ The data has a square data matrix.

◯ The data as a uniform distribution.

https://stats.idre.ucla.edu/spss/seminars/efa-spss/ (https://stats.idre.ucla.edu/spss/seminars/efa-spss/)

◀ 📄 Unit 8: Graded Assignment Quiz Score: **8** out of 8

Unit 8: Quiz

Due Apr 7 at 11:59pm   **Points** 8   **Questions** 8
Available until Apr 8 at 2:59am   **Time Limit** 45 Minutes

This quiz was locked Apr 8 at 2:59am.

Attempt History

	Attempt	**Time**	**Score**
LATEST	**Attempt 1**	5 minutes	8 out of 8

Score for this quiz: **8** out of 8
Submitted Apr 7 at 6:39pm
This attempt took 5 minutes.

Question 1	**1 / 1 pts**

Which of the following statements is incorrect for the PCA?

○ The principal components identify the direction where the spread of the
data is maximized.

○ The PCA algorithm finds all principal components simultaneously through
matrix multiplication.

Correct!   ◉ The PCA algorithm is based on iterative gradient search.

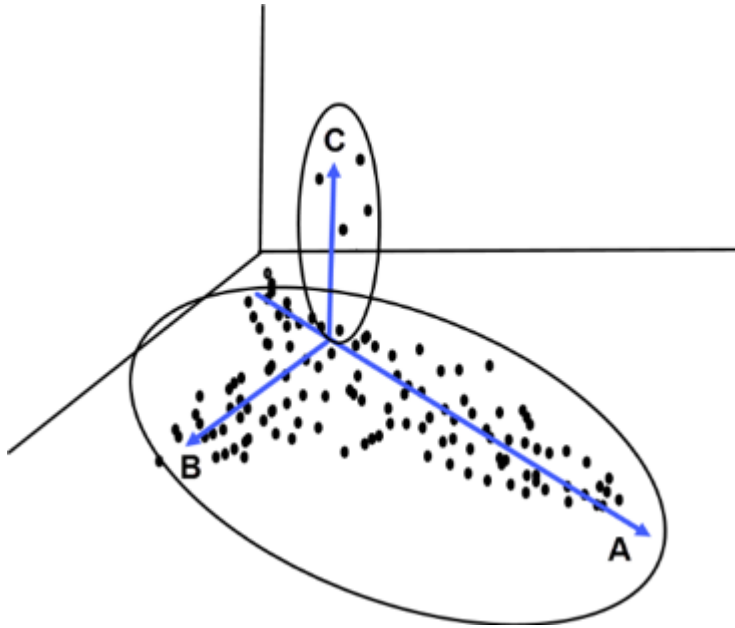○ The principal components are the linear projection of the original inputs.

Question 2

1 / 1 pts

In the following figure, which vector is the best candidate for the third principal component?

Correct!

- ⦿ Vector C

- ◯ Vector A

- ◯ Vector B

Question 3

1 / 1 pts

The most interpretable principal components are those that have the smallest eigenvalues.

- ◯ True

Correct!

- ⦿ False

Question 4

1 / 1 pts

Which of the following is NOT true for dimensionality reduction?

    ○ Dimensionality reduction is useful for data visualization.

Correct!

    ◉ Dimensionality reduction slows down the training of the machine learning systems.

    ○ Dimensionality reduction mitigates the curse of dimensionality.

    ○ Reducing dimensionality does lose some information so it may make machine learning perform slightly worse.

Question 5

1 / 1 pts

The first principal component for a dataset is found to be on the direction given by vector U = (4,2). Which of the following might be the direction of the second principal component? (Select all that apply.)

Correct!

    ☑ V=(-3, 6)

Correct!

    ☑ V=(-2,4)

    ☐ V=(-3,1.5)

    ☐ V=(1.5,3)

The principal components are orthogonal to each other.

Question 6

1 / 1 pts

Which of the following is WRONG for the PCA?

○ The PCA reduces the dimension of feature space.

Correct!

◉ The length of the first principal component vector is larger than that of the second principal vector.

○ The length of each principal component vector is equal to 1.

○ Any two principal components are orthogonal to each other.

Question 7

1 / 1 pts

Which of the following is one way to find out whether the dimensionality reduction algorithm performs well?

○ The dimension of the data matrix.

○ The number of principal components.

Correct!

◉ The measure of the reconstruction error.

○ The length of the principal component vectors.

Question 8   1 / 1 pts

Which of the following is an advantage of dimensionality reduction? (Select all that apply.)

Correct!

☑ Saving storage space

Correct!

☑

Visualizing the data and gaining insights on the most important features.

Correct!

☑ Improve model accuracy by removing redundant features and noise.

Correct!

☑ Less training time for a learning algorithm using the features.

Quiz Score: **8** out of 8

Mini-Assignment: Key Techniques for Deep Learning

Due Apr 16 at 11:59pm   **Points** 8   **Questions** 4
Available until Apr 17 at 2:59am   **Time Limit** 40 Minutes
Allowed Attempts 2

This quiz was locked Apr 17 at 2:59am.

Attempt History

	Attempt	Time	Score
LATEST	**Attempt 1**	1 minute	8 out of 8

Score for this attempt: **8** out of 8
Submitted Apr 16 at 10:47pm
This attempt took 1 minute.

Question 1   2 / 2 pts

Which option below describes the process of learning in the context of a fixed neural network architecture?

Correct!

- ⦿ Adapting the weights in response to different input output pairs

- ◯ Adapting the input to fit the desired output

- ◯ Manually updating the various weights

- ◯ Finding new connections in the neural network architecture

---

Question 2   2 / 2 pts

What does a neuron compute?

○ An activation function followed by a linear function (z = Wx + b)

Correct!   ◉ A linear function followed by an activation function

○ Only the activation function

○

The mean and standard deviation of input features followed by an activation function

Question 3   2 / 2 pts

In a 2-D convolutional neural network (CNN), what does the number of kernels define?

○ The number of output values

○ The number of input pixels

Correct!   ◉ The number of feature maps

○ The number of layers

Question 4   2 / 2 pts

_____ is an example of an unsupervised neural network, while _____ should be used if the input is a sequence in time.

○ Recurrent Neural Network, Autoencoder

○ Recurrent Neural Network, Convolutional Neural Network

○ Convolutional Neural Network, Autoencoder

Correct!

◉ Autoencoder, Recurrent Neural Network

Quiz Score: **8** out of 8

Unit 9: Quiz

Due Apr 16 at 11:59pm   **Points** 6   **Questions** 6

Available until Apr 17 at 2:59am   **Time Limit** 60 Minutes

This quiz was locked Apr 17 at 2:59am.

Attempt History

	Attempt	Time	Score
LATEST	**Attempt 1**	1 minute	5 out of 6

Score for this quiz: **5** out of 6
Submitted Apr 16 at 10:48pm
This attempt took 1 minute.

Question 1   1 / 1 pts

Which of the following created the first image classification application that classified images with at least 80% accuracy?

   ○ Oxford

   ○ MSRA

   ○ Google

Correct!   ◉ AlexNet

> Correct! AlexNet created the first image classification application that classified images with at least 80% accuracy.
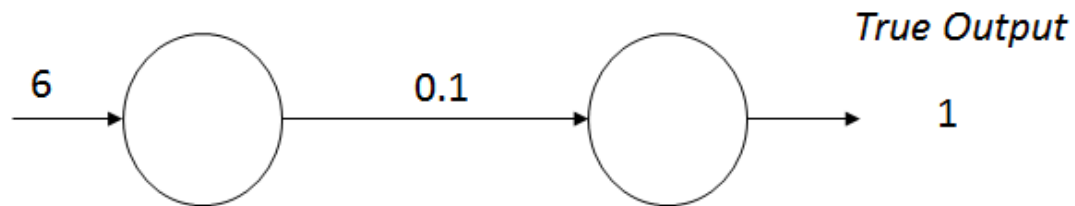

Question 2   1 / 1 pts

The figure below shows a simple neural network. An observation with one variable is presented to the network as shown. What is the updated weight if linear function y=x is used as the activation function and the Mean Squared Error as the error function?

$$Mean\ Squared\ Error\ =(y-\hat{y})^2$$

- ○ 0.16

Correct!

- ⦿ 1.06

  Correct! The updated weight from the neural network is 1.06.

- ○ 1.28

- ○ 0.38

Question 3

1 / 1 pts

Which of the following is true for the Perceptron neural network architecture?

Correct!

- ⦿ Each neuron is connected to all of the inputs.

  Correct! In the Perceptron neural network architecture, each neuron is connected to all of the inputs.

- ○ A bias feature is added to the output neuron only.

○ The output error is always less than 0.5.

○ The output is always greater than zero.

Question 4

0 / 1 pts

The figure below shows a simple neural network. An observation with two variables (0.8, 0.2) is presented to the network as shown. What is the predicted output from the neural network using a tanh activation function?

$$tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$
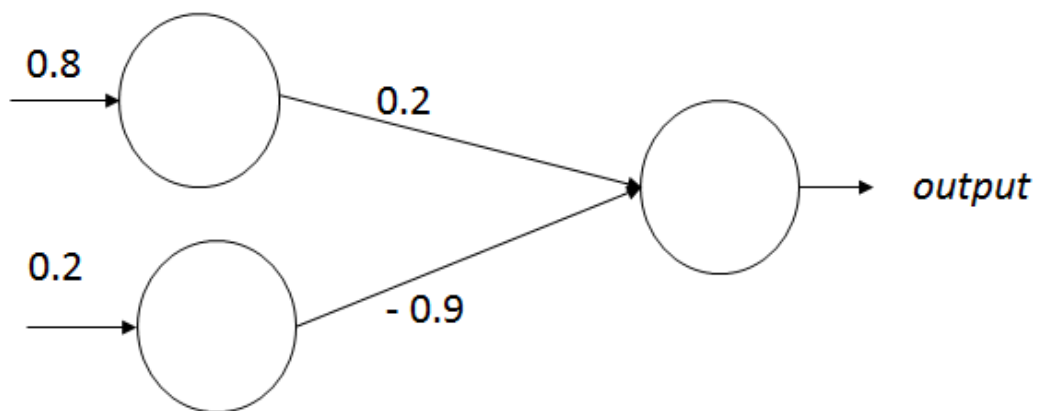
○ $\frac{e^{0.2} - e^{-0.2}}{e^{0.2} + e^{-0.2}}$

○ $\frac{e^{-0.2} - e^{-0.2}}{e^{-0.2} + e^{-0.2}}$

ou Answered

◉ $\frac{e^{-0.02} - e^{-0.02}}{e^{0.02} + e^{-0.02}}$

Incorrect. The predicted output from the neural network using a tanh activation function is not $\frac{e^{-0.02} - e^{-0.02}}{e^{0.02} + e^{-0.02}}$. Calculate the value of $\sum_i w_i x_i$. Then use the tanh function.

orrect Answer   ○ $\frac{e^{-0.02}-e^{0.02}}{e^{-0.02}+e^{0.02}}$

Question 5

1 / 1 pts

Which of the following is NOT true for pooling in CNNs?

☐ It is a method of reducing the number of features for the next layer.

Correct! ☑ It is a method for increasing the number of features for the next layer.

> Correct! The purpose of pooling is not to increase the number of features.

☐ It is a method of feature extraction.

☐ It is a method of summarizing neighboring feature detectors.

Question 6

1 / 1 pts

How many pooling layers does LeNet architecture have?

○ 4

Correct! ◉ 2

○ 6

○ 8

Quiz Score: **5** out of 6