Foundation of ML Quiz - 4

Duration of Quiz is 30 minutes (strict), 8 PM to 8:30 PM. You can resubmit until 8:30 PM. No negative marks, but questions may carry unequal marks.

The respondent's email (**cs18btech11001@iith.ac.in**) was recorded on submission of this form.

Program (PhD, MTech, MDS, BTech) *

- (●) BTech
- ( ) MTech
- ( ) MDS
- ( ) PhD

Name *

Abburi Venkata Sai Mahesh

Course ID (FoML course ID you have registered in AIMS : CS5590, AI5000, SM5000, AI2000) *

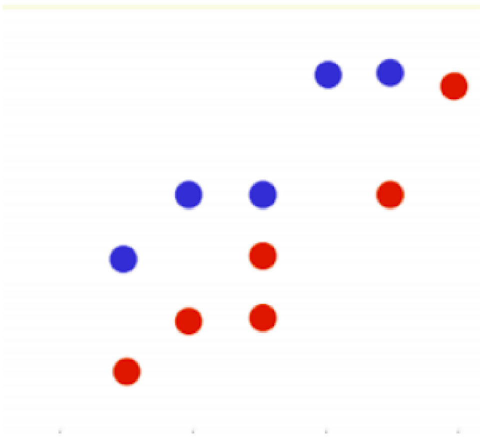
- (●) CS5590
- ( ) AI5000
- ( ) SM5000
- ( ) AI2000

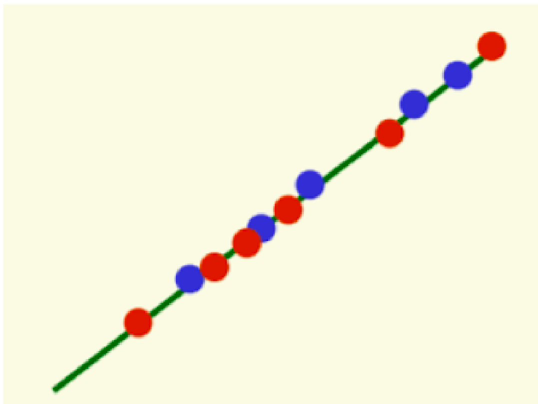
Roll no *

CS18BTECH11001

Quiz begins here ....

✓ [True/False] Consider the following data , with positive examples labeled as blue and negative examples labeled as red. The figure on the right side represents the Fisher linear discriminant projection. [Marks : 2]





- ○ True
- ◉ False ✓

✓ Which of the following does not indicate $A \perp B|C$? $A \perp B|C$ means $p(A,B|C) = p(A|C)p(B|C)$ [Marks 2]

- ○ $A \rightarrow C \rightarrow B$
- ○ $A \leftarrow C \leftarrow B$
- ○ $A \leftarrow C \rightarrow B$
- ◉ $A \rightarrow C \leftarrow B$ ✓

✓ [True/False] In kernel PCA, assume S proportional to the sample covariance matrix in the feature space, and K is the kernel matrix over data points. L is an eigenvalue of S if and only if L is also an eigenvalue of K. [2 Marks]

- (●) True ✓
- ( ) Option 2

✗ [True/False] Fisher linear discriminant maximizes wS1w, where S1 is [2 Marks]

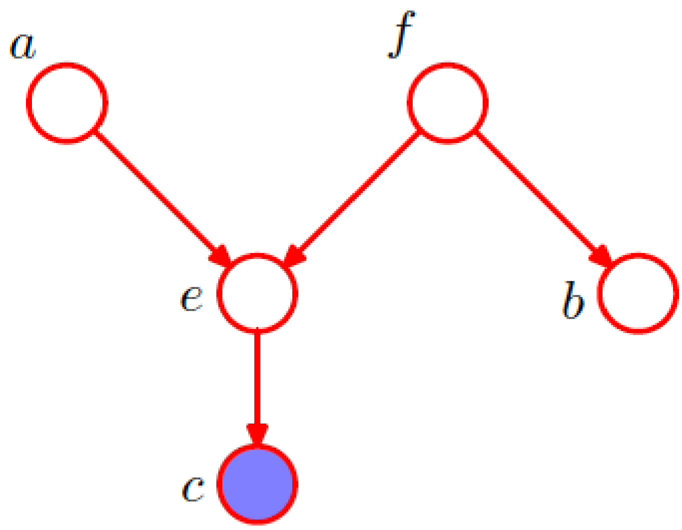
$$\mathbf{S}_1 = \sum_t (\mathbf{x}^t - \mathbf{m}_1)(\mathbf{x}^t - \mathbf{m}_1)^T r^t$$

- (●) True ✗
- ( ) False

✗ Assume for a classification data, within class scatter is given by Sw and between class scatter is given by Sb. Then Fisher's linear discriminant find [Mark : 2]

- ( ) largest eigen vectors of Sw^{-1} * Sb
- ( ) largest eigen vectors of Sb^{-1} * Sw
- ( ) largest eigen vectors of Sw * Sb
- ( ) largest eigen vectors of Sw
- ( ) largest eigen vectors of Sb

Consider the probabilistic graphical model below. Which is the correct statement ? [Marks : 3]



- [x] a and b are independent
- [ ] a and b are not independent

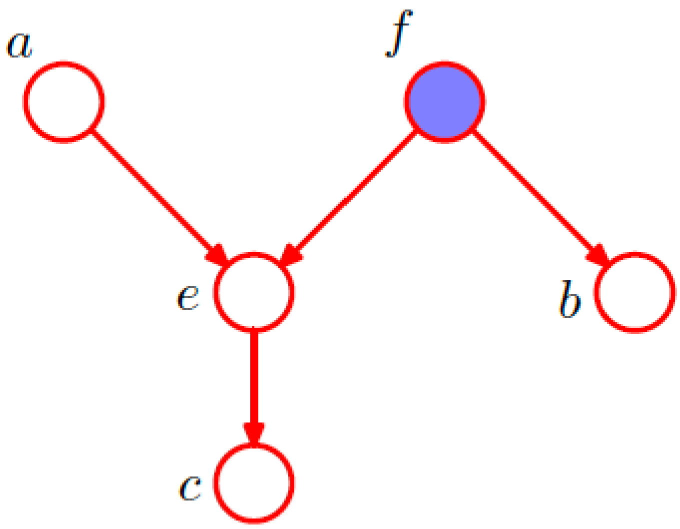
[True/False] Given a design matrix X = R^{n * d}, where d << n (n samples, d dimension, d is much less than n), if we project our data onto a k dimensional subspace using PCA where k equals the rank of X, we recreate a perfect representation of our data with no loss. [2 Marks]

- [ ] True
- [ ] False

[True/False] In probabilistic PCA, marginal distribution p(x) cannot be obtained in a closed form [2 Marks]

- [ ] True
- [x] False

Consider the probabilistic graphical model below. Which is the correct statement ? [Marks : 3]

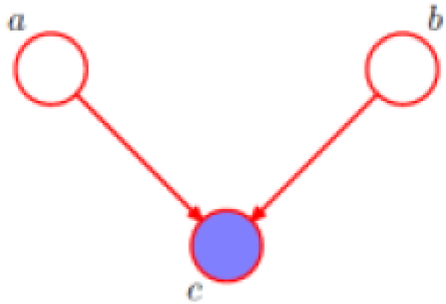


- a and b are independent
- a and b are not independent

[True/False] which of the following statements on Hidden markov model is correct ? [Marks : 2]

- HMM is a discriminative markov random field
- HMM is a generative Bayesian network
- HMM is a discriminative Bayesian network
- HMM is a generative markov random field

✓ [True/False] In the following probabilistic graphical model, a and b are independent. [Marks : 2]



- ◯ True
- ◉ False ✓

✓ Which of the following statements is true? [2 Marks]

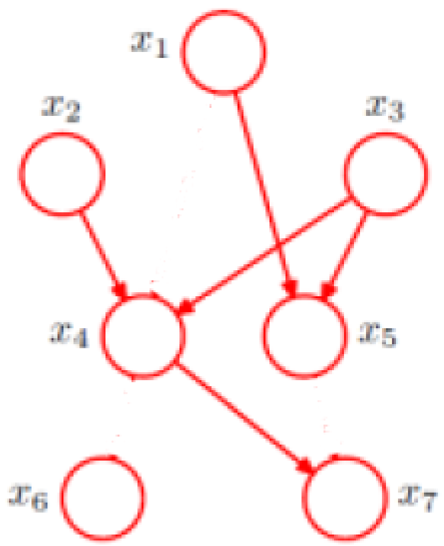
$$\text{let } A = \begin{bmatrix} 0 & 2 & -1 \\ -2 & 8 & -5 \\ -3 & 10 & -7 \end{bmatrix}. \text{ Then } X = \begin{bmatrix} 0 \\ -1 \\ -2 \end{bmatrix}, Y = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}.$$

- ◯ X is eigenvector of A with the eigenvalue 2.
- ◯ Y is eigenvector of A with the eigenvalue 2.
- ◯ X and Y are eigenvectors of A with the eigenvalues 2,-2.
- ◉ X is eigenvector of A with the eigenvalue -2. ✓

Assume that we have a training set of fixed size N and that all features are uniformly distributed on [0; 1]. Predicted response y correspond to the average of the responses associated to the training examples that are near x. Now suppose that we wish to make a prediction of a test example x by creating a d-dimensional hypercube centered around x that contains on average 10% of the training examples. Assuming dimension d=10, what is the length of each side of the hypercube? [Marks :3]

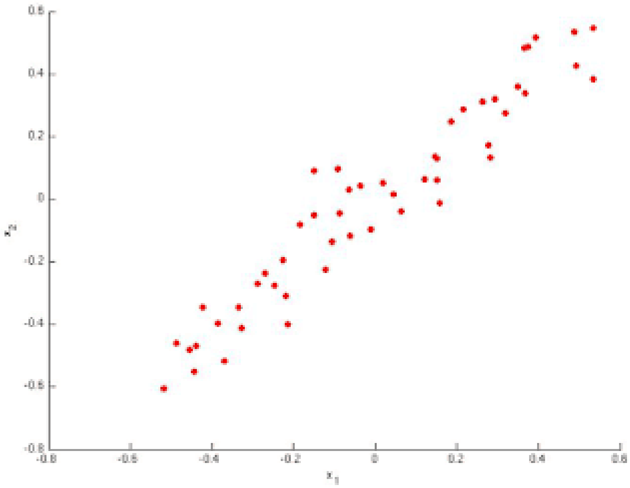
- 1
- 0.1
- 0
- 0.8
- 0.3

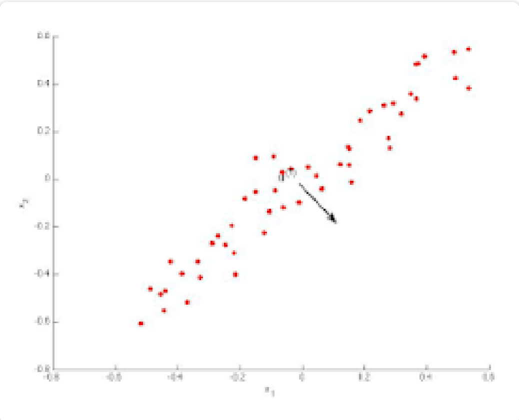
Whats the joint distribution modeled through the following probabilistic graphical model ? [Marks : 3]



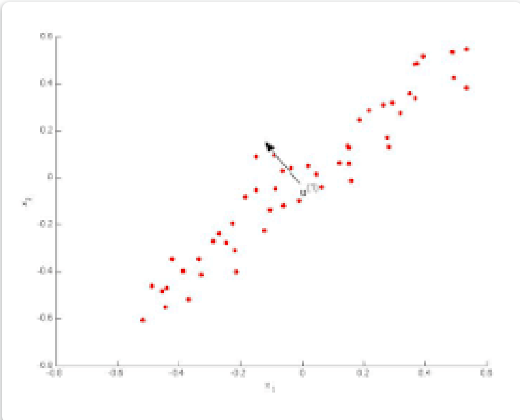
- p(x4|x2)p(x3)p(x5|x1)p(x1)p(x7|x4)p(x2)p(x6)
- p(x1|x5)p(x2|x4)p(x3|x4,x5)p(x4|x7)p(x5)p(x7)p(x6)
- p(x2|x4)p(x3|x4,x5)p(x4|x7)p(x1|x5)
- p(x3)p(x1)p(x5|x1,x3)p(x7|x4)p(x6)p(x2)p(x4|x2,x3)
- p(x3)p(x4|x2,x3)p(x1)p(x5|x1,x3)p(x7|x4)p(x2)

Consider the following 2 dimensional data set. Which of the following figures correspond to possible values that PCA may return for the first eigen vector / first principal component? [Marks 2]

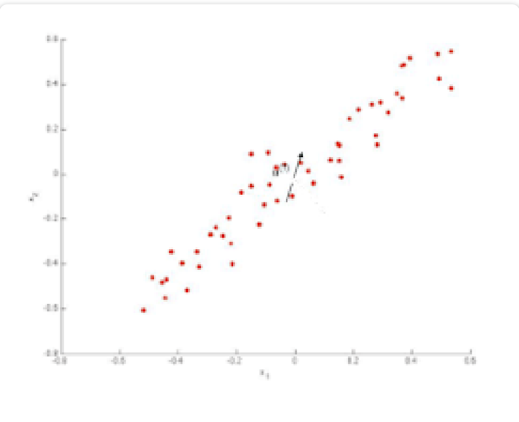




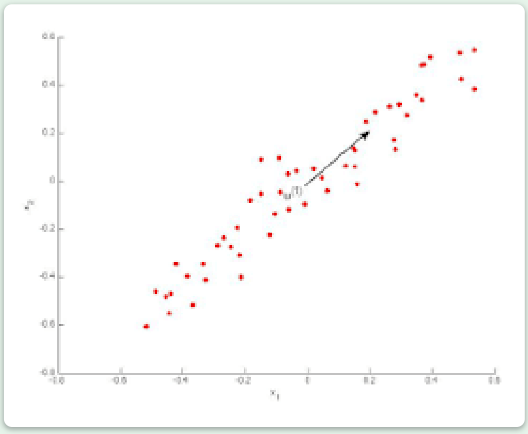
Option 1



Option 2



Option 3



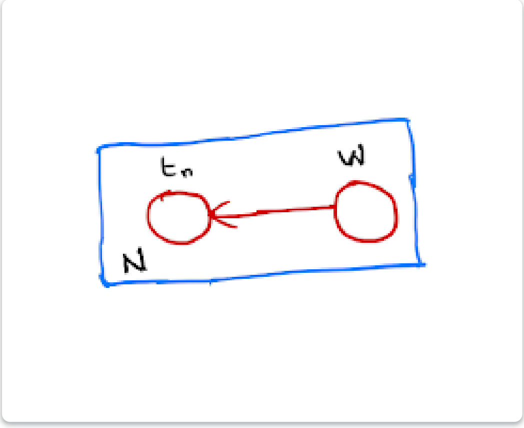
Option 4

Assume that we have a training set of fixed size N and that all features are uniformly distributed on [0; 1]. Predicted response y correspond to the average of the responses associated to the training examples that are near x. Suppose that we have 100 features (d = 100) and we want to predict using only training examples that are within 10% of the input range in both dimensions. On average, what fraction of the training examples will we use to make each prediction? [Marks :2]

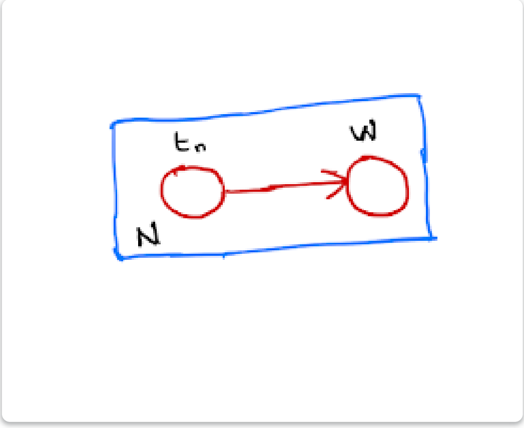
- 100%
- 10%
- 1%
- 0%

Which of the following is the correct plate notation for the joint distribution below [Marks : 3]

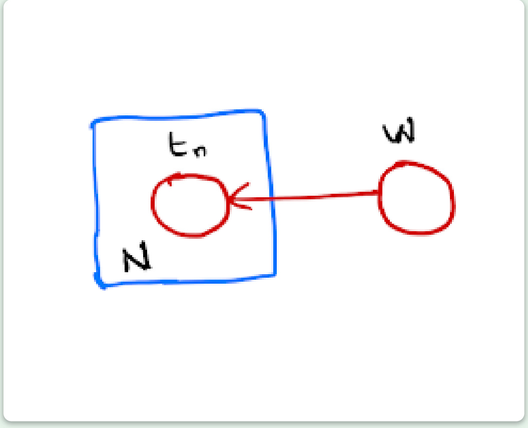
$$p(\mathbf{t}, \mathbf{w}|\mathbf{x}, \alpha, \sigma^2) = p(\mathbf{w}|\alpha) \prod_{n=1}^{N} p(t_n|\mathbf{w}, x_n, \sigma^2).$$



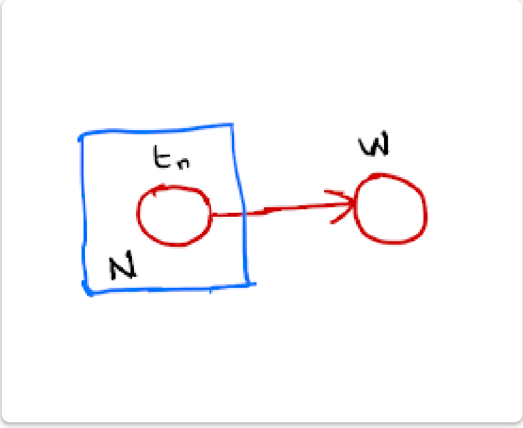
Option 2



Option 3



Option 1



Option 4

For a sequence of inputs [x1,x2,x3] and sequence of outputs [y1,y2,y3], HMM assumes p(x1,x2,x3,y1,y2,y3) can be written as [Marks : 3]

- (selected, incorrect) p(y1|x1)p(x2|x1)p(y2|x2)p(x3|x2)p(y3|x3)
- p(x1|y1)p(y2|y1)p(x2|y2)p(y3|y2)p(x3|y3)
- p(x1|y1)p(x2|x1)p(x2|y2)p(x3|x2)p(x3|y3)
- p(y1|x1)p(y2|y1)p(y2|x2)p(y3|y2)p(y3|x3)

Let A be a real, symmetric n x n matrix. Which of the following are true about A's eigenvectors and eigenvalues? [1 Mark]

- (selected, correct) A can have no more than n distinct eigenvalues
- The vector 0 is an eigenvector, because A0 = lambda 0
- A can have no more than 2n distinct unit-length eigenvectors
- We can find 2n mutually orthogonal eigenvectors of A

Given N number of d-dimensional data [Xi] (i=1...N), you run principle component analysis and pick P principle components. Can you always reconstruct any data point [Xi] for i=1..N, from the P principle components with zero reconstruction error? [2 Marks]

- Yes, if P < d
- (selected, correct) Yes, if P = d
- Yes, if P < N
- No always

Why is PCA sometimes used as a preprocessing step before regression? [1 Marks]

- To reduce over fitting by removing poorly predictive dimensions.
- To make computation faster by reducing the dimensionality of the data.
- To expose information missing from the input data.
- For inference and scientific discovery, we prefer features that are not axis-aligned.

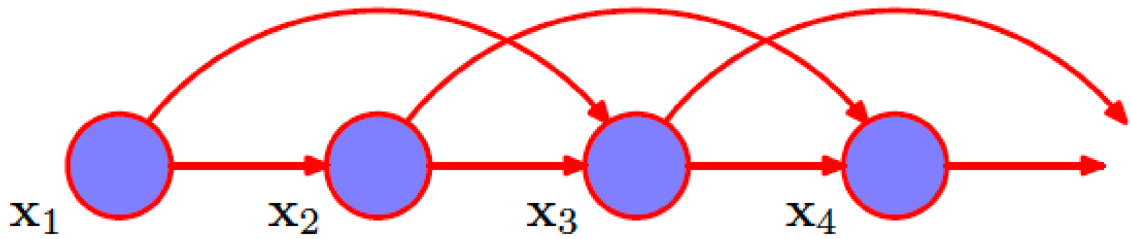
Probabilistic PCA models latent representation 'z' as a ______ while GMM models latent variable 'z' as a ______ [2 Marks]

- Multinomial, Gaussian
- Multinomial, Gamma
- Gaussian, Gaussian
- Gamma, Multinomial
- Gaussian, Multinomial

[True/False] As the number of dimensions increases, the percentage of the volume in the unit ball shell with thickness 'eps' grows. [1 Marks]

- True
- False

Consider the graphical model below following a second order markov assumption. Suppose the observations are discrete variables taking 5 different values. The number of parameters required to represent the model is [Marks : 3]



- 100
- 625
- 20
- 25

Assume that we have a training set of fixed size N and that all features are uniformly distributed on [0; 1]. Predicted response y correspond to the average of the responses associated to the training examples that are near x. Suppose that we have three features (d = 3) and we want to predict using only training examples that are within 10% of the input range in both dimensions. On average, what fraction of the training examples will we use to make each prediction? [Marks :2]

- 10%
- 5%
- 1%
- 0.1%

You are given a design matrix X. .Let's use PCA to reduce the dimension from 2 to 1. The unit eigenvectors, and the corresponding eigenvalues, of the covariance matrix of the design matrix X are A1 with eigen value 2 and A2 with eigen value 162. What one-dimensional subspace are we projecting onto? For each of the four sample points in X (not the centered version of X), write the coordinate (in principal coordinate space, not in R2) that the point is projected to. [Marks : 4]

$$X = \begin{bmatrix} 6 & -4 \\ -3 & 5 \\ -2 & 6 \\ 7 & -3 \end{bmatrix} \qquad \text{A1} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \qquad \text{A2} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} \end{bmatrix}$$

$(6,-4)\to\frac{10}{\sqrt{2}}$, $(-3,5)\to\frac{8}{\sqrt{2}}$, $(-2,6)\to\frac{8}{\sqrt{2}}$, $(7,-3)\to\frac{10}{\sqrt{2}}$

- [ ] Option 1

$(6,-4)\to\frac{2}{\sqrt{2}}$, $(-3,5)\to\frac{2}{\sqrt{2}}$, $(-2,6)\to\frac{4}{\sqrt{2}}$, $(7,-3)\to\frac{4}{\sqrt{2}}$

- [ ] Option 2

$(6,-4)\to\frac{10}{\sqrt{2}}$, $(-3,5)\to\frac{-8}{\sqrt{2}}$, $(-2,6)\to\frac{-8}{\sqrt{2}}$, $(7,-3)\to\frac{10}{\sqrt{2}}$

- [x] Option 3 ✓

$(6,-4)\to\frac{-10}{\sqrt{2}}$, $(-3,5)\to\frac{8}{\sqrt{2}}$, $(-2,6)\to\frac{8}{\sqrt{2}}$, $(7,-3)\to\frac{-10}{\sqrt{2}}$

- [ ] Option 4

- [ ] Other:

This form was created inside of IIT Hyderabad.

Google Forms