Assignment 8

The due date for submitting this assignment has passed. Due on 2018-09-26, 23:59 IST.

As per our records you have not submitted this assignment. 

1) PCA is closely related to _________: indeed, some statistical packages deliberately conflate 1 point the two techniques.

- Linear regression
- Factor analysis
- Eigenvalue
- SAS

No, the answer is incorrect.
Score: 0

Accepted Answers:
Linear regression

2) Which is the correct statement: 1 point

- PCA is sensitive to local optima so multiple random initializations should be done
- If you do not perform mean normalization, PCA will rotate data in some random way which is not desired
- PCA can reduce the dimensionality of data only by 1
- All of the above

No, the answer is incorrect.
Score: 0

Accepted Answers:
If you do not perform mean normalization, PCA will rotate data in some random way which is not desired

3) Which of the following is the best way to select the number of principal components k? (n = dimensionality of input data; m = number of input examples) 1 point

- Use elbow method
- Choose k to be 99% of m
- Choose k that minimizes approximation error = \( \frac{1}{m} \sum (x_i - x_i\text{-approximate}) \)
- Choose k to be the smallest such that ≥ 99% of the variance is retained

No, the answer is incorrect.
Score: 0

Accepted Answers:
If you do not perform mean normalization, PCA will rotate data in some random way which is not desired

A project of
© 2014 NPTEL - Privacy & Terms - Honor Code - FAQs - In association with

Funded by

© 2014 NPTEL - Privacy & Terms - Honor Code - FAQs - In association with

Funded by
No, the answer is incorrect.
Score: 0
Accepted Answers:
Choose k to be the smallest such that ≥ 99% of the variance is retained

4) The key concept of factor analysis is that multiple observed variables have similar patterns of responses because they are all associated with a latent (i.e. not directly measured) variable.  1 point

TRUE
FALSE

No, the answer is incorrect.
Score: 0
Accepted Answers:
TRUE

5) If there are n observations with p variables, then the number of distinct principal components is:  1 point

- min(n-1,p)
- min(n,p)
- max(n-1,p)
- max(n-1,p)

No, the answer is incorrect.
Score: 0
Accepted Answers:
min(n-1,p)

6) Suppose in a principal component analysis you get following eigenvalues \( \lambda_1, \lambda_2, \ldots, \lambda_p \). What is the proportion of variance explained by including only principal component 1  1 point

- \( \frac{\lambda_1}{\lambda_1 + \lambda_2 + \ldots + \lambda_p} \)
- \( \lambda_1 \)
- \( \frac{\lambda_1}{\lambda_1} \)
- \( \frac{\lambda_1 + \lambda_2 + \ldots + \lambda_p}{1} \)

No, the answer is incorrect.
Score: 0
Accepted Answers:
\( \frac{\lambda_1}{\lambda_1 + \lambda_2 + \ldots + \lambda_p} \)

7) The __________ represents the total variance explained by each factor.  1 point

- Eigenvalue
- Correlation matrix
- Bartlett’s test of sphericity
- Residuals

No, the answer is incorrect.
Score: 0
Accepted Answers:
Eigenvalue

8) Factor analysis is a __________ and __________ technique 1 point

- Data Analysis and Decision Making - I - - Unit 11... https://onlinecourses.nptel.ac.in/noc18_mg33/un...
9) A principal components analysis was run and the following eigenvalue results were obtained: 2.831, 2.118, .542, .441, .283, .085. How many factors would you retain using the eigenvalues to determine the number of factors

- 1 point
- 4
- 6

No, the answer is incorrect.
Score: 0
Accepted Answers:
- data reduction and summarization

10) For a real square matrix $A$, if there is a real number $\lambda$ and a non-zero vector $x$ such that $Ax = \lambda x$, then $\lambda$ is called an ________ and $x$ is called an ________

- Eigenvalue, eigenvector
- Eigenvector, eigenvalue
- Component, eigenvector
- Dimensionality, eigenvector

No, the answer is incorrect.
Score: 0
Accepted Answers:
- Eigenvalue, eigenvector

11) PCA involves the calculation of the eigenvalue decomposition of a data ________ or singular value decomposition of a data matrix, usually after mean centering the data for each attribute.

- Random variable
- Covariance matrix
- Multivariate normal distribution
- File

No, the answer is incorrect.
Score: 0
Accepted Answers:
- Covariance matrix

12) In a vector space, for any vector $v$ and a unit vector $u$, we have $v = v \cdot u + (v - v \cdot u u)$. Finding the vector $u$ that minimizes $| (v - v \cdot u u) |$ is the same as finding a vector $u$ that maximizes $|v \cdot u|$. 1 point
PCA computes the direction \( u \) (principal component) that maximizes the variance of the data along the direction \( u \).

PCA computes the direction \( u \) (principal component) that minimizes the variance of the data along the direction \( u \).

PCA computes the direction \( u \) (principal component) that sums up the variance of the data along the direction \( u \).

PCA calculates the direction \( u \) (principal component) that minimizes the variance of data from the direction \( u \).

No, the answer is incorrect.
Score: 0

Accepted Answers:

**PCA computes the direction** \( u \) (principal component) **that maximizes the variance of the data along the direction** \( u \).

13 Choose the correct statement:  

- In linear regression, the distance between observed point \((x_i,y_i)\) from the computed point \((x_i,a+bx)\) on line \(y=a+bx\), is minimized along the \( y \) direction. In PCA, the distance of the observed point \((x_i,y_i)\) is minimized to a line which is orthogonal to the line \(y=a+bx\), is minimized
- Linear regression and PCA are same techniques
- Linear regression and PCA can only be applied when there is multicolinearity issue
- In linear regression, the distance between observed point \((x_i,y_i)\) from the computed point \((x_i,a+bx)\) on line \(y=a+bx\), is maximized along the \( y \) direction. In PCA, the distance of the observed point \((x_i,y_i)\) is maximized to a line which is orthogonal to the line \(y=a+bx\), is minimized

No, the answer is incorrect.
Score: 0

Accepted Answers:

*In linear regression, the distance between observed point \((x_i,y_i)\) from the computed point \((x_i,a+bx)\) on line \(y=a+bx\), is minimized along the \( y \) direction. In PCA, the distance of the observed point \((x_i,y_i)\) is minimized to a line which is orthogonal to the line \(y=a+bx\), is minimized*

**Factor analysis is related to principal component analysis (PCA), but the two are not identical.**

- TRUE
- FALSE

No, the answer is incorrect.
Score: 0

Accepted Answers:

TRUE

**Any multivariate joint distribution can be written in terms of univariate marginal-distribution functions and a ______ which describes the dependence structure between the variables.**

- Copula
- Principal component
- Covariance matrix
- Random variable

No, the answer is incorrect.
Score: 0
Accepted Answers:
*Copula*