**Question Bank Module 6**

1. Using PCA reduce the dimensions from 2 to 1

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Features | Sample 1 | Sample 2 | Sample 3 | Sample 4 |
| X | 4 | 8 | 13 | 7 |
| Y | 11 | 4 | 5 | 14 |

1. Using PCA reduce the dimensions from 2 to 1

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Features | Sample 1 | Sample 2 | Sample 3 | Sample 4 | Sample 5 | Sample 6 |
| X (symbolic BP) | 126 | 128 | 128 | 130 | 130 | 132 |
| Y (diastolic BP) | 78 | 80 | 82 | 82 | 84 | 86 |

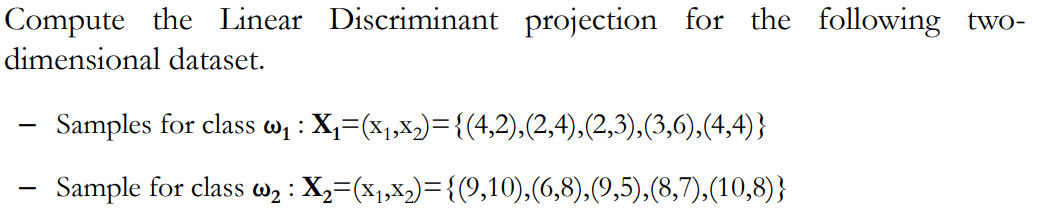
Linear discriminant analysis

Q1. Compute the linear discriminant projection for the following two-dimensional set (10 Marks)

C1=((1,2) (2,3) (3,3) (4,5) (5,5))

C2=((1,0) (2,1) (3,1) (3,2) (5,3) (6,5))

Q2. (10 Marks)



Q3. What is curse of dimensionality (2 marks)

Q4. Explain the concept of factor analysis with practical example

Concept Video: <https://www.youtube.com/watch?v=24KBj9Oo9Zk>

<https://www.youtube.com/watch?v=gexvMJacB0g>

Q5. Explain the concept of multidimensional scaling (5 Marks)

Concept Video: <https://www.youtube.com/watch?v=7_Wt2kwZdGU>

Q6. What is feature mapping? What are the different methods use for feature mapping?

Ans: Feature mapping is a technique used in data analysis and machine learning to transform input data from a lower-dimensional space to a higher-dimensional space, where it can be more easily analyzed or classified. Feature mapping involves selecting or designing a set of functions that map the original data to a new set of features that better capture the underlying patterns in the data. The resulting feature space can then be used as input to a machine learning algorithm or other analysis technique. Feature mapping can be used in a wide range of applications, from natural language processing to computer vision, and is a powerful tool for transforming data into a format that can be analyzed more easily. However, there are also potential issues to consider, such as the curse of dimensionality, overfitting, and computational complexity.

Feature mapping, also known as feature engineering, is the process of transforming raw input data into a set of meaningful features that can be used by a machine learning algorithm. Feature mapping is an important step in machine learning, as the quality of the features can have a significant impact on the performance of the algorithm.

There are several techniques for feature mapping, including:

Feature extraction: This involves transforming the input data into a new set of features that capture the most important information. For example, in image processing, features such as edges, corners, and textures can be extracted from the image.

Feature transformation: This involves applying mathematical functions to the input data to transform it into a new set of features. For example, in text classification, the input text can be transformed into a bag-of-words representation, where each word in the text is represented by a count of its frequency.

Feature selection: This involves selecting a subset of the available features that are most relevant to the task at hand. This can be done using techniques such as mutual information, correlation, or regularization.

Feature scaling: This involves scaling the features to ensure that they have similar ranges and are on the same scale. This is important for some algorithms, such as those based on distance metrics.

Feature mapping can be a time-consuming and iterative process, as different feature combinations and transformations may need to be tried to find the best set of features for a given task. However, effective feature mapping can significantly improve the performance of a machine learning algorithm and enable it to make more accurate predictions.

Feature engineering: This is the process of creating new features from the existing ones in order to improve the performance of a machine learning algorithm. Feature engineering can involve a combination of feature extraction, transformation, and selection techniques. For example, in a fraud detection task, a new feature could be created by calculating the ratio of the transaction amount to the average transaction amount for that user.

Dimensionality reduction: This involves reducing the number of features in a dataset while retaining the most important information. This can be done using techniques such as principal component analysis (PCA) or t-distributed stochastic neighbor embedding (t-SNE). Dimensionality reduction can help to reduce overfitting and improve the computational efficiency of a machine learning algorithm.

Embeddings: An embedding is a vector representation of a feature or object that captures its semantic meaning. For example, in natural language processing, word embeddings can be used to represent words as dense vectors in a high-dimensional space, where words with similar meanings are closer together. Embeddings can be learned using techniques such as word2vec or GloVe.

Data augmentation: This involves creating new examples by applying transformations to the existing data. For example, in image classification, data augmentation can involve rotating, flipping, or cropping the images to create new variations. Data augmentation can help to increase the size of the training set and improve the robustness of a machine learning model.

Domain knowledge: Expert knowledge about the domain in which the machine learning algorithm will be applied can be used to guide the feature mapping process. For example, in a medical diagnosis task, domain knowledge about the symptoms and risk factors of a particular disease can be used to select or engineer relevant features.