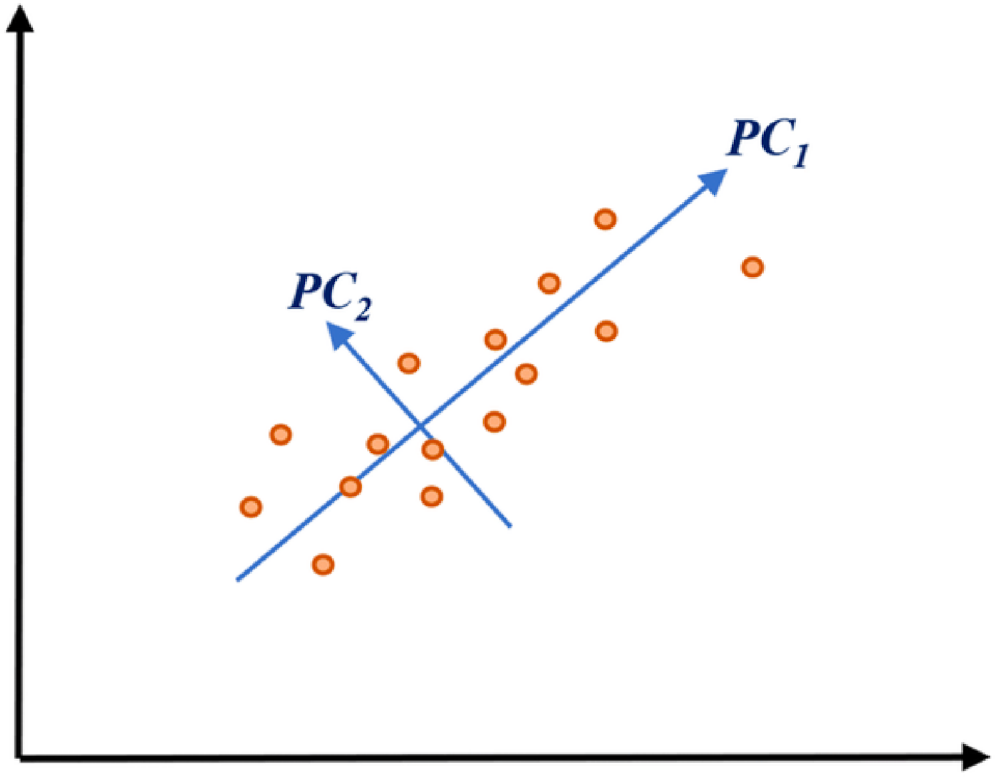
Day 19: Dimensionality Reduction



Author: Alisha Metkari

Dimensionality Reduction

- The number of input features present in data are called dimensions.
- Higher dimensional data increases time complexity of the model.
- Reducing dimensions is called the dimensionality reduction.
- Some datasets contain huge number of features which increases the redundancy and model complexity.
- To tackle this, there are dimensionality reduction techniques.
- These techniques help to reduce higher dimensional data into lower dimensional data, providing similar information.

Author: Alisha Metkari

Curse of Dimensionality Reduction

- Handling large dimensional data is very difficult, it is called curse of dimensionality.
- As the number of features increases the model becomes more complex.
- Complex models take more time in training and prediction.
- Curse of dimensionality is not good in machine learning. We need to tackle this before actually building the model.

Author: Alisha Metkari

Methods of Dimensionality Reduction

- There are two methods for dimensionality reduction: feature selection and feature extraction.
- Feature selection is the process of selecting valuable features from complete set which represents most of the information of complete data. It basically avoids the redundant features. We have discussed this approach in previous post.
- Feature extraction is the process of transforming higher dimensional data into lower dimensional data with equivalent amount of information. It keeps less number of features with the most amount of information. Some common feature extraction methods are Principal Component Analysis(PCA), Kernel PCA.

Author: Alisha Metkari

Principal Component Analysis (PCA)

- Principal Component Analysis (PCA) is a statistical process that converts the correlated features into a set of linearly uncorrelated features with the help of orthogonal transformation.
- It is an unsupervised method. It uses mathematical concept called Singular Values Decomposition (SVD).
- PCA identifies a set of orthogonal axes, called principal components, that capture the maximum variance in the data. The principal components are linear combinations of the original variables in the dataset and are ordered in decreasing order of importance. The total variance captured by all the principal components is equal to the total variance in the original dataset.

Author: Alisha Metkari

Principal Component Analysis (PCA)

- The first principal component captures the most variation in the data, but the second principal component captures the maximum variance that is orthogonal to the first principal component, and so on.
- PCA assumes that the higher the variation in a feature, the more information that features carries.
- It can be used for data visualisation, feature extraction, data compression.

Author: Alisha Metkari

Kernel PCA

- The PCA we discussed previously is the simplest form of the PCA which is linear transformation and works on linear data.
- It is not suitable for non-linear data.
- To deal with non-linear data, we use kernel trick.
- A kernel trick is a method to project non-linear data onto a higher dimensional space and separate different distributions of data.
- Kernel trick uses kernel function to map non-linear data into linear space.
- Once the distributions are separated we can use PCA to separate them linearly.

Author: Alisha Metkari

Advantages

- Dimensionality reduction reduces the space complexity.
- It reduces the time complexity.
- It removes the redundant features, making the model perform well.

Author: Alisha Metkari