KOMINFO digitalent

PRO Professional Academy

Digital Talent Scholarship 2022

Math For Machine Learning - PCA 3

Lead a sprint through the Machine Learning Track

KOMINFO digitalent

Agenda

- Statistics of Datasets
- Inner Products

KOMINFO digitalent

Are your students ML-ready?

KOMINFO digitalent

Recap

Learning Objectives

- Summarize PCA
- Write code that implements PCA
- Assess the properties of PCA when applying to high-dimensional data

What is PCA?

- Principal Component Analysis (PCA) is one of the most fundamental dimensionality reduction techniques that are used in machine learning.
- PCA reduced the dimensionality of the data by projecting them into a lower-dimensional subspace (called principal subspace)
- To apply PCA we will use some concepts that we already learned such as eigenstuff, variance, and covariance

Principal Component Analysis

KOMINFO digitalent

DEMO : PCA from scratch

When we should use PCA ?

- If you want to reduce the dimensionality of the data but can't identify which variables to completely remove from consideration
- If you are comfortable making your data & model less interpretable
- If you want to ensure your variables are independent of one another

Step in PCA

- Standardize the high-dimensional data.
- Calculate covariance matrix.
- Calculate eigenvector & eigenvalue of the covariance matrix.
- Choose the eigenvectors associated with the M largest eigenvalues to be the basis of the principal subspace.
- Orthogonal projection of the data onto the principal axis.

KOMINFO digitalent

[StatQuest: Principal Component Analysis (PCA), Step-by-Step - YouTube]

DEMO : Supervised Learning using PCA

KOMINFO digitalent

Unsupervised Learning using PCA

Impact of using PCA

- Speed up the training process by reducing the dimensionality of the data
- Improve the accuracy of the classification model

KOMINFO digitalent

Q & A

KOMINFO digitalent

Thank You