06/02/2019

MSc. Data Science
Project Proposal

Group name (optional):
Group members : Arkaprava Sinha, Subhasish Basak, Purnendu Ghosh, Rohan Khaitan

---

- **Title:** Principal Component Analysis and its application in Machine Learning
- **Description:** While analysing real world datasets, it is very difficult to work with a large number of features. When faced with a large set of correlated variables, principal components allow us to summarize this set with a smaller number of representative variables that collectively explain most of the variability in the original set.
- **Motivation:** Some benefits of dimensionality reduction include less computation/training time and removing multicollinearity by eliminating redundant features. It also helps in visualising multidimensional data.
- **Project details:**
  - The objective of this project is to analyse how PCA can be used to find patterns to reduce the dimensions of the dataset with minimal loss of information.
  - **Tools**: Singular Value Decomposition, Eigen-values & Eigen vectors, Python
  - **Outcome**: Implementing Principal Component Analysis in unsupervised learning using python.
  - **Plan**:
    We will describe how Numerical Linear Algebra is applied in Principal Component Analysis.
    We will implement PCA in a real world dataset.
- **References:**

Introduction to Statistical Learning using R, Gareth James et al
Numerical Linear algebra, Lloyd N.Trefethen, David Bau II
