**FoDS Assignment #1**

**Group Members**

1. Akshit Khanna 2017A7PS0023P
2. Vitthal Bhandari 2017A7PS0136P

**Task #1 Curse of Dimensionality**

Identifying and demonstrating the problems due to the curse of dimensionality in high dimensional data. Preliminary problems identified :-

1. Overfitting of Data (With limited samples)
2. Increased Computations on Data (Demonstrated by showing comparison of time for classification on different number of dimensions)
3. Concentration effect of Lp-Norms (Demonstrated using histogram showing distribution of distances between all samples)
4. Relevant and Irrelevant features (Demonstrated with sample data generated)

Preliminary Solutions to CoD are PCA for linear dimensionality reduction and Kernel PCA for nonlinear dimensionality reduction.

Standard datasets in R will be used to demonstrate the advantage of dimensionality reduction. For standard PCA the datasets used will be **Sonar, Mines vs. Rocks (208 instances, 61 attributes)** and **Johns Hopkins University Ionosphere database (351 instances, 35 attributes).** For Kernel PCA we would use moons or concentric circles dataset to demonstrate the advantage of Kernel PCA over PCA.

**Task #2 Gradient Descent (GD)**

Identifying the problems in gradient descent and implementing basic gradient descent on a regression task in R. Preliminary problems identified :-

1. It can lead to local minima for a non-convex loss function
2. Choice of a suitable learning rate(A large value will cause it to overshoot before reaching the minima while a small value will require large number of epochs)
3. Slow for large datasets(Large number of epochs required)

Preliminary Solutions for the above problems in standard gradient descent are solved using variants of gradient descent such as stochastic gradient descent and batch gradient descent and can be combined with momentum for further improvement. **Boston Housing (506 instances, 14 attributes)** dataset will be used for testing regression using gradient descent.

**Work Division**

Identification of Problems with illustrations **(Vitthal Bhandari)**, Implementation of PCA and Kernel PCA **(Akshit Khanna)**, Implementation of GD **(Akshit Khanna)**, Implementation of variations of GD **(Vitthal Bhandari).**

All the implementations will be done in **R**.