**Lab 4 – Dimensionality Reduction and MLP**

***Group members:***

1. *Duc Phi Ngo (Student ID: 616922)*
2. *Huynh Anh Vu Nguyen (Student ID: 618069)*
3. *Thanh Do Nguyen (Student ID: 615941)*
4. *Van Ty Pham (Student ID: 616938)*

1. Give some intuitive reason why unit hypersphere volume decreases (rather than increases) as of dimension exceeds 5.

Let's start with a unit circle in 2D. If we add a third dimension, we have two ways to extend it:

* Make a cylinder by shifting the circle up and down. Its volume is π × r² × h = 3.14 × 1 × 2 = 6.28.
* Make a sphere by rotating the circle. Its volume is 4.18, which is less than the cylinder.

The reason the sphere has less volume is that rotation covers less space than a direct shift. Instead of extending in a straight line (like a cylinder), the sphere spreads out, reducing its volume.

Now, if we keep adding dimensions, the same effect happens: the hypersphere covers less space than its surrounding hypercube. That’s why the volume of a hypersphere shrinks in higher dimensions. This idea also matches the mathematical formula for hypersphere volume.

2. Dimensionality Reduction is very important, especially, considering using Machine Learning in Big Data applications.  Below are the 2 papers you need to read.  These will help better understand what we covered in the class. These would also help you in your project as well as when you will join the industry.   The first paper is relative easier than the 2nd one which has more math.

a. <https://ieeexplore.ieee.org/document/9036908>

**Analysis of Dimensionality Reduction Techniques on Big Data**

b. <http://jmlr.org/papers/volume16/cunningham15a/cunningham15a.pdf>

**Linear Dimensionality Reduction:  
Survey, Insights, and Generalizations**

Just write a short summary of the **first paper based on your own understanding.**

First paper explains how Dimensionality Reduction (DR) helps in handling big data by making machine learning models faster and more accurate. It describes different DR techniques like:

* Principal Component Analysis (PCA) – Finds important patterns in data and removes unnecessary features.
* Linear Discriminant Analysis (LDA) – Helps in classification by separating different categories in data.
* t-SNE – Used for visualizing high-dimensional data in 2D or 3D.

- Autoencoders – A deep learning method that compresses data while keeping important information.

The paper highlights that big data often has too many features, which can make models slow and inefficient. DR helps by removing redundancy, reducing computation time, and improving model performance. It is especially useful in machine learning, deep learning, and data visualization.

3. **Write any key points that you have learned from the paper that you can possibly use in your project.**  In case you are not addressing any DR for your project now, write how DR can affect your project in future when you may have large data with many dimensions.

In future projects, when data grows large and complex, Dimensionality Reduction will be crucial to ensure efficiency, accuracy, and interpretability. How DR Can Affect a Machine Learning Project with Large Data and Many Dimensions:

* Speeds Up Computation – Reduces the number of features, making model training faster.
* Avoids the Curse of Dimensionality – Improves generalization by reducing sparsity in high-dimensional space.
* Improves Model Accuracy – Removes redundant and irrelevant features, reducing noise.
* Enhances Visualization – Helps in understanding complex datasets by projecting them into 2D/3D.
* Reduces Overfitting – Prevents models from memorizing unnecessary details, improving performance.