

## BMEG3105 Lecture 9 scribing (Ma Wing Laam)

- What is Classification?
- Classification involves categorizing records from a training set based on their attributes, with the goal of assigning classes to unseen records.
- How to do Classification:
  - 1. Training Data: Use existing records with known classes.
  - 2. Classification Method: Apply a method to determine the class for new data.
  - 3. Data to Classify: New instances are classified based on learned attributes.
- K-Nearest Neighbors (KNN)
  - Basic Idea: Classify an instance based on the majority class of its K nearest neighbors.
  - Training Process: Store available training instances.
  - Predicting Process:
    - 1. Determine the K closest training instances to the query.
    - 2. Return the most frequent class label among those instances.
  - Normalization: Data should be normalized before classification.
  - Choosing K in KNN
    - A typical value for K is between 5 and 10, which works well for most low-dimensional datasets.
    - Cross-validation can help in determining a suitable K.
  - KNN example
    - 1. Normalize the data.
    - 2. Compute distances between the query instance and training instances.
    - 3. Identify the K most similar instances.
    - 4. Determine the mode class from those instances.
- Clustering vs. Classification
  - Clustering: Groups data based on similarity without predefined classes.
  - Classification: Assigns known classes to new data based on a training set.
- Unsupervised Learning: Analyzes unlabelled data (clustering)
- Supervised Learning: Involves labeled data for training (classification)