CMPSC445 Applied Machine Learning Exam 1 Review Outline

Exam 1 covers the contents taught from modules 1 to 5 which include basic concepts, algorithms, and applications on machine learning as listed below (it is not necessarily exhaustive).

1. Explain machine learning concepts and algorithms:

- The main characteristics of machine learning and the relation to data science.
- Machine learning categories: Supervised learning versus unsupervised learning, the main ideas and their flow charts.
- Data feature (attribute), dimension, label
- Reasons for separating training data and test data; Reasons for performing cross-validation in training;
- Supervised learning algorithms
    - Linear regression
    - Logistic regression
    - Decision tree
    - KNN
    - Naive Bayes
    - SVM
- Unsupervised learning algorithm: K means;
- Curse of dimensionality; PCA for dimensionality reduction;
- Optimization technique: Gradient descent
- Explain factors that affect accuracy in machine learning:
  sample size, curse of dimensionality, over/under fitting, hyper-parameter tuning etc.
- Comparison of machine learning methods.

2. Apply Python programming and its libraries in solving machine learning problems.

- Python programming basics and data structures such as list.
- Understand Python libraries (modules, packages).
- Be able to use Scikit-learn to create the ML pipeline we learned so far:
  Data preparation->define model->train model->evaluation and validation->test->predictions
- Know what the following are used for:
    - Numpy
    - Pandas
    - Matplotlib