IMT 574 Midterm Theory

1. There is a lot around us that's driven by some form of machine learning (ML), but not all. Give an example of a system/service that does not use ML (1 point) and the one that does (1 point). Use this contrast to explain ML in your own words (2 points). [4 points]

   Take traditional retail POS and cloud-based retail POS as examples. The former only store sales data on local database, and it cannot provide deep analysis of customer behaviors. However, the cloud-based system can use the previous sales data to recommend customers products, provide more predictions for stores. Thus, ML is a machine can learn from previous experience, and use this ability to help humans finish some tasks.

2. When/why would you use unsupervised learning? [1 point]

   Do not have labels / identify inherent structures -> usually use in exploratory analysis and dimensionality reduction

3. Many of the ML models are represented using parameters. Use this idea to define ML. [2 points]

   Parameters are essential parts of ML models. They can define what kind of skills the models can "learn". Models then need parameters to make predictions.
   For example, in a linear regression model y = mx + c, m and c are parameters for this model.

4. When would you use batch gradient descent (BGD) (1 point) and when would you use stochastic gradient descent (SGD) (1 point) ? Why (2 points)? [4 points]

   BGD: the number of training examples is small -> use all the training data
   SGD: the number of training examples is large -> use just one example of training data at a time

5. Decision tree is a greedy algorithm. List one advantage (1 point) and one disadvantage (1 point) of being greedy here. [2 points]

   +: fast, perform well on large datasets
   -: does not take into account the global optimum; overfitting

6. How does Random Forest address the issue of bias or overfitting? [2 points]

   use a random subset of features
   train on different samples of the data

7. Calculate the entropy of seeing a six by rolling a fair die. Show calculations. [6 points]

$$Entropy(S)_{\,binary} = -p_i * log_2(p_i) - (1-p_i) * log_2(1-p_i)$$
$$= -\frac{1}{6} * log_2\left(\frac{1}{6}\right) - \frac{5}{6} * log_2\left(\frac{5}{6}\right) = 0.65$$
$$Entropy(S) = -\sum_{i=1}^{k} p_i * log_b(p_i)$$
$$(1) = -\sum_{i=1}^{k} p_i * log_6(p_i) = -\frac{1}{6} * log_6(\frac{1}{6}) * 6 = 1$$
$$(2) = -\sum_{i=1}^{k} p_i * log_2(p_i) = -\frac{1}{6} * log_2(\frac{1}{6}) * 6 = 2.585\ (-1\ point)$$

8. What ML technique and/or algorithm (e.g., classification using random forest) would you use to solve the following problems? [4 points] (1 point for each question)

- Predicting rainfall in inches using various weather condition signals:
  linear regression
- Trying to model different seasons on Mars without knowing which day represents which season:
  K-mean clustering or other unsupervised learning approaches
- Declaring if it will snow or not on Christmas:
  Logistic regression or other classification approaches
- Figuring out the importance (weight) of diet and exercise in one's well-being:
  Linear regression

9. Compare cost function in gradient descent and likelihood function in gradient ascent. [2 points]

gradient descent is a **linear function** (0.5 points) to find a local minimum and the hypothesis here is continuous
gradient ascent is a **sigmoid function** (0.5 points) to find a local maximum and the hypothesis here is discrete

10. You are trying to figure out (1) if temperature on the election day has any association with individuals voting or not (1 point), and (2) the overall turnout rate (1 point). And if there is that association, you want to make prediction model(s) (1 point). What techniques/algorithms would you use for each of these? [3 points]

(1) classification - decision tree
(2) regression - linear regression
(3) any supervised learning models - linear regression