PEKING UNIVERSITY, GUANGHUA SCHOOL OF MANAGEMENT
Machine Learning and Artificial Intelligence

Assignment 1

Prof. Yingjie Zhang

Due date: October 17, 2021 11:59pm

Instruction

- This homework includes both conceptual questions (described in this .pdf file) and implementation questions (described in a separate .ipynb file)
- Deliverables: Please submit **one pdf file** (with all your answers to the conceptual questions) and **your IPython notebook files** (including one with a .ipynb extension and one with a .html extension) that include all coding work (with necessary comments).
- You may discuss the questions with fellow students, however you must always write your own solutions and must acknowledge whom you discussed the question with. Do not copy from other sources, share your work with others, or search for solutions on the web. Plagiarism will be penalized according to university rules.

1 General ML Questions

1.1 Identify a Learning Problem

[7 points] Describe the movie recommendation learning problem by stating as precisely as possible the *task*, *performance measure*, and *training experience*.

1.2 Machine Learning Types

[8 points] For each task below, specify what type of machine learning problem it is (i.e., supervised or unsupervised; classification or regressions or others). Explain your reasoning briefly in 1-2 sentences each.

- Predict the stock market index

- Identify whether or not an alumni is going to donate to PKU

- Recommend online courses that are better taken together

- Segment customers based on social-demographic attributes

2 Decision Tree

Suppose we want to predict whether a flight on a particular day will be canceled based on three factors: whether there is a snowstorm, where it is an official holiday, and whether it is a long-distance (≥ 4 hours) flight. Table 2.1 shows the training examples. Please answer the following three questions.

No.	Snowstorm	Holiday	Long Distance	Canceled
1	T	F	T	No
2	T	F	T	Yes
3	F	F	T	No
4	T	F	T	No
4	F	F	F	No
6	F	F	F	Yes
7	T	F	F	Yes
8	F	F	F	Yes

Table 2.1: Flight Data Training Examples

1. If we chose "Holiday" as the root of a decision tree, what would be the effects? Explain in terms of information gain. [4 points]

2. If "Holiday" is not proper as the root node, which attribute will you choose as the root node of the decision tree? Explain your reasons with necessary calculations. [6 points]

3. Describe your complete decision tree in words (sample format: If there is a snowstorm, the flight will be canceled). [5 points]