IE406 Midterm Exam (Take-home)

Due Date: 11:59 PM, April 22nd (Friday)

1. Please explain how existing machine learning techniques can be applied for different applications below in one paragraph each.

(1) Self-driving cars (Autonomous cars) (5 points)

(2) Illegal parking detection (5 points)

(3) Credit card fraud detection (5 points)

(4) Spam e-mail classification (5 points)

(5) Product display in retail store (5 points)

2. A machine learning-based car safety system, which detects whether or not the car crashes, has been developed based on the data using 20 different car sensors and has been trained using a large volume of historical data. It was discovered that the software in some sensors was wrong after developing and training the classifier, so the software has been updated to correct errors. Please explain what consequences you can expect for this machine learning-based system and what you can do about them. (10 points)

3. Suppose that we have a classification task with categories A, B, C, …

(1) Please find an application(s) where mistakes are more critical for some categories than others. For example, please explain a situation(s) that to classify a test instance as A, where the correct answer is B, is more dangerous than vice versa. (5 points)

(2) Please explain whether or not the issue above has effects on your training process and evaluation protocol. (5 points)

4. Explain the difference between traditional machine learning and deep learning. (5 points)

5. Explain the difference between DBSCAN and K-means. (2 points)

6. What is the Apriori principle? (Hint: You can use the terms "frequent", "itemset", and/or "subsets" if they are necessary.) (3 points)

7. Explain the difference between classification and regression. (5 points)

8. Consider a transactional database where A, B, C, D, E, F, and G are items. Let the minimum support be 50% and minimum confidence be 70%. Please show all frequent itemsets and all association rules. (15 points)

TID	Items
1	A, B, C, E
2	A, B, C, D, E
3	A, B, C, G
4	A, C, F
5	A, B, D, E, F

9. Suppose that you are requested to mine association rules for 1,000,000,000 transactions. Please suggest an approach that speeds up association rule mining, using sampling. (5 points)

10. Consider the following table containing three binary input A, B, and C and a binary output Y. Suppose you have to predict Y using naïve Bayes.

A	B	C	Y
1	0	0	0
0	1	1	0

0	0	1	0
1	0	0	1
0	0	1	1
0	1	0	1
1	1	0	1

(a) Predict the probability P(Y=0 | A=0, B=1, C=1) without Laplace smoothing (5 points)

(b) Predict the probability P(Y=0 | A=1) without Laplace smoothing (5 points)

(c) Predict the probability P(Y=0 | A=0, B=1, C=1) with Laplace smoothing (5 points)

(d) Predict the probability P(Y=0 | A=1) with Laplace smoothing (5 points)