First Examination 2024
Artificial Intelligence and Machine Learning

You are required to submit the following files through email. 1. A pdf file containing your solutions and report for every problem. The file name should be "report.pdf". In the report file you should follow the format that has been introduced in the class. 2. A pdf file containing your source code for every problem. The file name should be "source.pdf". 3. The necessary testing files requested in every problem. The file name should be as requested in every problem. **You should upload all files to your account in Google Drive no later than 14:00 pm, Oct. 19, 2024. Any files submitted later than the due time will not be accepted.**

Please be reminded that this is a formal exam. You are required not to discuss with anyone and not to release any part of your solutions to others. Anyone who fails to follow the rule will be required to leave this school. Please respect yourself and your classmates.

1. Given a dataset in the file named trn1.csv for training, please train a maximum likelihood estimator to learn a classifier for the data with 8 classes. Test your classifier on the testing dataset in the file named tst1.csv. (1) Please calculate training accuracy and the discriminant functions $g_i(\boldsymbol{x})$, $i = 1 \ldots 8$. (2) Please calculate the determinant of covariance matrix in $g_i(\boldsymbol{x})$, $i = 1 \ldots 8$. In other words, please calculate $|\Sigma_i| \equiv \det(\Sigma_i)$, $i = 1 \ldots 8$. (3) Please generate a file named result1.csv to contain the testing results based on the testing data in the file tst1.csv. Each row of the data in result1.csv should contain [testing data x], [class]. (25%)

2. Given a dataset in the file trn2.csv for training, please train a maximum likelihood estimator to learn a regression model $f(\boldsymbol{x})$ for the data. You may use a 10-th order linear polynomial to approximate the model $f(\boldsymbol{x})$. In other words, you may use the polynomial $g(\boldsymbol{x}) = \boldsymbol{w}^T \boldsymbol{x} + w_0$, $\boldsymbol{w} = [w_1, w_2, ..., w_{10}]^T$, $\boldsymbol{x} = [x_1, x_2, ..., x_{10}]^T$, to approximate $f(\boldsymbol{x})$. Test your regression model on the testing dataset in the file named tst2.csv. Please (1) show the regression model g($\boldsymbol{x}$) based on your training. (2) calculate the coefficient of determination of your model. (3) generate a file named result2.csv to contain the testing results based on the testing data in the file tst2.csv. Each row of the data in result2.csv should contain [data $\boldsymbol{x}$], [$g(\boldsymbol{x})$]. (25%)

3. Use the training data in problem 1 to do classification again. (1) Use a common covariance matrix to train your maximum likelihood estimator. Please calculate the determinant of common covariance matrix and your $g_i(\boldsymbol{x})$, $i = 1 \ldots 8$. Generate a file named result3_1.csv to contain the testing results based on the testing data tst1.csv. (2) Use a common diagonal matrix to train your maximum likelihood estimator. Please calculate the determinant of the common diagonal

covariance matrix. Please show your $g_i(\boldsymbol{x})$, $i = 1 \ldots 8$. Generate a file named result3_2.csv to contain the testing results based on the testing data in tst1.csv. (25%)

4. Given a dataset named apriori_data.csv, please design an Apriori algorithm to learn the association rules with minimum support 0.05 and minimum confidence 0.6. Generate a file named result4.csv to contain all the association rules and the confidence associated with every rule. Please describe the details how you generate these association rules. Each row of the data should follow the format:
[(item_a)'^'(item_b)'^'…'^'(item_h)'||'(item_w)'^'(item_x)'^'…'^'(item_z)], [confidence]
(25%)

The exam is due at 14:00 pm, Oct. 19, 2024.