University No:_____

THE UNIVERSITY OF HONG KONG
FACULTY OF ENGINEERING
DEPARTMENT OF COMPUTER SCIENCE
COMP7409 Machine Learning in Trading and Finance

Date: May 13, 2022 Time: 6:30pm-8:30pm

INSTRUCTIONS:

a. Candidates are permitted to refer to any printed/handwritten materials in the examination. Internet searching and crowdsourcing from group messages, online forums or social media, etc. are strictly forbidden.

b. You should hand-write your answers on self-prepared blank papers. Typing and writing answers on digital devices are NOT accepted.

c. This paper has 4 questions. Answer ALL questions.

d. Total mark is 100.

e. Write your university number clearly at the beginning of your answer script. DO NOT write down your name.

f. At the end of the examination, submit to OLEX a single pdf file containing your answer script.

g. Only approved calculators as announced by the Examinations Secretary can be used in the examination. It is the candidates' responsibility to ensure that their calculator operates satisfactorily, and candidates must record the name and type of the calculator used on the front page of your answer script.

University No:_____

All questions carry equal marks.

1. Consider the problem of training a machine learning model for a credit card company. The model should be a binary classifier that suggests how to respond to a new potential customer, or more precisely, suggests either a positive response, which is to "issue a credit card to the customer", or a negative response, which is to "decline a credit card". There are two kinds of mistakes that the model can make:
    - false positive, which suggests issuing a credit card to someone who will become a bad credit risk, and
    - false negative, which declines a potentially good credit card holder.

a. Which is the more serious type of errors, a false positive or a false negative? Justify your answer.

Suppose that we have trained a neural network N for this credit card problem, and the output of the model is a positive real number between 0 and 1. To use N to solve the problem, we need to determine a threshold t between 0 and 1 such that for any customer, if the number returned by N for this customer is greater than t, we give a positive response; otherwise, we give a negative one. We denote this model of neural network N with threshold t as (N,t).

To measure the performance of (N, t) on a set of customers, we can compute its *sensitivity* and *specificity*, which are defined as follows:

Sensitivity of (N,t) = TP / (TP + FN)

Specificity of (N,t) = FP / (TN + FP)

were TP (True Positive) is the total number of customers that (N, t) gives a positive response correctly, and FP (False Positive) is the total number of customers that (N, t) gives a positive response incorrectly. Similarly, TN (True Negative) is the number of customers that (N,t) gives a negative response correctly, and FN (False Negative) is the number of those (N,t) gives a negative response incorrectly.

b. Suppose there are 172 customers, 126 of them are positive customers and should be given positive responses, i.e., we should issue credit cards to them, and the remaining 46 customers should be given negative responses. If the outputs of the neural network N are summarized in the following table, compute the sensitivity and specificity of (N, t) = (N, 0.65) for this set of customers. (To simplify the question, we assume that N's outputs are only in {0.1.,0.2, ... 1.0})

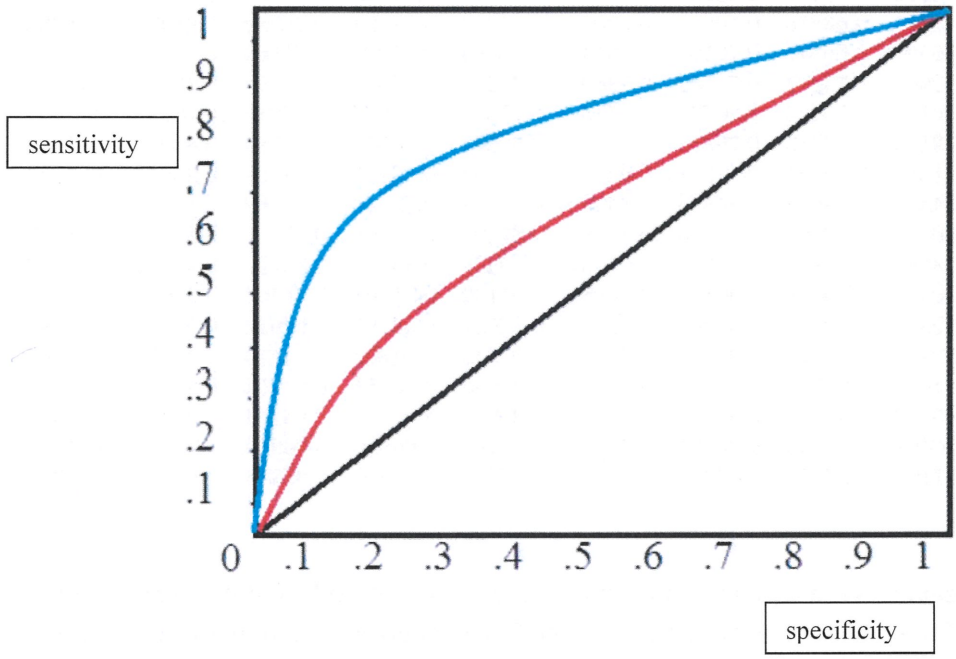
outputs	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Positive customers	2	7	5	10	8	20	28	25	18	3
Negative customers	5	9	7	8	4	3	5	2	3	0

c. Propose a model (N, t) that gives the best sensitivity (i.e., as large as possible).

d. Propose a model (N, t) that gives the best specificity (i.e., as small as possible).

University No:_____

e. Consider any neural network N. Note that if we increase the threshold t from 0 to 1, the sensitivity and specificity of (N, t) will change accordingly and the pair (specificity, sensitivity) will trace a curve on a 2-D plane. We call such a curve the ROC curve for N. The following diagram shows three ROCs for three different neural networks for solving our credit card problem on a set of customers, and let N_{blue} be the neural network giving the blue ROC curve, N_{red} giving the red curve, and N_{black} giving the black curve. If you need to pick one from these three neural networks to solve the credit card problem, which one should you pick? Justify your answer.



University No:_____

2. a. Consider the following call option for a particular stock in which the current stock price is $340, the strike price is $300, and it will expire after 55 days. Assume that the volatility rate is 18%, risk-free rate is 4%. Determine the price of this option using the Black-Stroke-Merton equation given in the lecture. You may use in your calculation the following table for the Cumulative Distribution Function of the normal distribution N(.).

x	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0.5	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.5279	0.53188	0.53586
0.1	0.53983	0.5438	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.6293	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.6591	0.66276	0.6664	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.7054	0.70884	0.71226	0.71566	0.71904	0.7224
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.7549
0.7	0.75804	0.76115	0.76424	0.7673	0.77035	0.77337	0.77637	0.77935	0.7823	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.8665	0.86864	0.87076	0.87286	0.87493	0.87698	0.879	0.881	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.9032	0.9049	0.90658	0.90824	0.90988	0.91149	0.91308	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.9222	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.9452	0.9463	0.94738	0.94845	0.9495	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.9608	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.9732	0.97381	0.97441	0.975	0.97558	0.97615	0.9767

b. Recently many researchers are trying to apply machine learning methods to determine option prices. State some of the deficiencies of using the Black-Stroke-Merton equation for option pricing and why machine learning may be a better method.

University No:_____

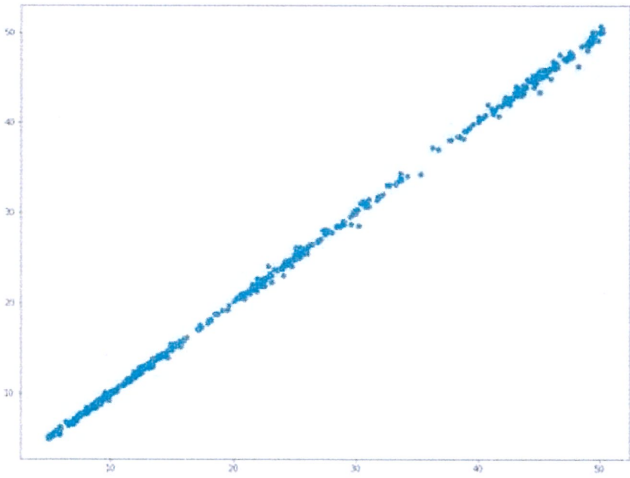
3. In Lecture 3, we have studied the following program for stock price prediction

```
tclose = shuffle(data['Today_close'], random_state=0)
nopen = shuffle(data['Nextday_open'], random_state=0)
X_train, X_test, y_train, y_test = train_test_split(tclose, nopen, test_size=0.2,
                                                    random_state=0)

X_trainp=X_train.to_numpy().reshape(-1,1)
X_testp=X_test.to_numpy().reshape(-1,1)

reg=linear_model.LinearRegression()
reg.fit(X_trainp, y_train)
```

and shown by the following diagram that the program's predictions are very accurate: the points (actual stock price, predicted stock price) are all falling closely to the line y = x.



a. Will you recommend your clients to use this program for stock investment? Why?
b. Suggest some simple idea(s) to improve the performance (i.e., to improve the probability that the program can help investors gain profits) of the program.

4. Compare the popularity of "Nike" and "Adidas" by writing a python program that computes, for everyday between May 7, 2022 to May 13, 2022, the average sentiment scores of related tweets posted on twitters on that day using the sentiment analysis tool "flair". You may assume the Bearer Token for accessing the tweets is stored at the first line of the file "BearerToken.txt". Your program should include all necessary libraries.

END OF PAPER