CITY UNIVERSITY OF HONG KONG

Course code & title	:	CS5489 Machine Learning: Algorithms & Applications
Session	:	Midterm, Semester A 2021
Time allowed	:	Two hours (Oct 28, 1:00pm-3:00pm)

This is the answer sheet for the CS5489 Midterm
Put all your answers in this document.

---

1. This paper consists of 13 questions.
2. Answer ALL questions.
3. Write your answers in the accompanying **"CS5489-midterm-2021A-answersheet.docx"**.

---

*This is an **open-book** final exam, see the allowed resources below.*

Instructions:

- Answer all questions in the accompanying Word document **"CS5489-midterm-2021A-answersheet.docx"**.
- The following resources are **allowed** during the final:
  - Videos of CS5489 lectures on Zoom,
  - any **unaltered** material on the CS5489 Canvas page, including lecture notes, tutorials, etc.
  - **unaltered** course textbooks
- Any other resources are **not allowed**, for example
  - internet searches.
  - Classmates.
  - other textbooks, other notes.
  - any text/notes copied into your lecture notes or textbook.
  - translation software.
- You should stay on Zoom during the entire exam time in case there are any announcements.
  - If you have any questions, please use the private chat function of Zoom to message Antoni.
- By 3:00pm Oct 28, submit the completed final to the "Final" Assignment on Canvas.
  - If you have trouble accessing Canvas, then you can send the completed docx via email to Antoni (abchan@cityu.edu.hk).

Below is the statement of academic honesty. Read it and put your Name, EID, and student ID to acknowledge that you agree with it and will follow its terms.

Statement of Academic Honesty

I pledge that the answers in this exam are my own and that I will not seek or obtain an unfair advantage in producing these answers. Specifically,

- ❖ *I will not plagiarize (copy without citation) from any source;*
- ❖ *I will not communicate or attempt to communicate with any other person during the exam; neither will I give or attempt to give assistance to another student taking the exam; and*
- ❖ *I will use only approved devices (e.g., calculators) and/or approved device models.*
- ❖ *I understand that any act of academic dishonesty can lead to disciplinary action.*

I pledge to follow the Rules on Academic Honesty and understand that violations may led to severe penalties.

Name: <KEY>
EID: <PUT YOUR EID HERE>
Student ID: <PUT YOUR STUDENT ID HERE>

Multiple Choice/Selection Questions (30 points)
5 marks each. *For a multiple selection question, an incorrect answer will be penalized 5/K marks, where K is the number of correct answers. If more incorrect answers are given than correct answers, the marks will be 0.*

Q1 ANSWER: <A, C, D>

Q2 ANSWER: <A, C, E >

Q3 ANSWER: <A, D, E >

Q4. ANSWER: <A, B, C >

Q5. ANSWER: <B, D >

Q6. ANSWER: <C, D, E >

Discussion Questions (70 marks)
10 marks each question.

Q7 ANSWER:

- [4 marks] The main problem of this situation is that the model is under-fitting the data, i.e., the model is not expressive enough to learn the features.
- [6 marks, 3 marks each] 2 Solutions
  - Change the feature representation: Introduce non-linear variables such as polynomial features, different feature encodings (one-hot encoding).
  - reduce regularization parameters;
  - use non-linear classifiers: change to non-linear classifiers such as kernel SVM, RF, Adaboost, etc.

Q8 ANSWER:

- [6 marks, 2 marks each] They differ on their general principles used for learning
  - SVM maximizes the margin (geometric principles).
  - LR maximizes the posterior probability of the data
  - NB is a generative model that learns the CCDs to fit the data.
- [4 marks] From the "No free lunch" theorem, there is no best classifier.

Q9 ANSWER:

- [2 marks] Choice: Random forest.
- [4 marks, 2 marks each] 2 Reasons
  - RF can reveal the important features.
  - RF can form non-linear decision boundary, thus can handle the non-linear relation between features and classes
  - RF is fast and usually brings good generalization..
- [2 marks] LR disadvantage: data may not be linearly separable.
- [2 marks] Kernel SVM disadvantage: too much data to store the kernel matrix.

Q10 ANSWER:

- [4 marks; 2 marks each] There are 2 issues:
  - 1) the data unbalance problem of the classifier (more negative examples vs positive examples)
  - 2) the classifier imbalance problem – as a screening test, we don't want to miss any positive samples. That is errors on the positive class (false negatives) are undesirable.
- [6 marks; 3 marks each] How to address:
  - 1) increase the weight on the positive class during training.
  - 2) increase the weight on the positive class during training AND adjust the classification threshold during test time.

Q11 ANSWER:

a) [2 marks] this is the L2 regularization term that prevents overfitting, by keeping the weights from becoming too large.

b) [2 marks] alpha is determined through cross-validation procedure.

c) [6 marks] Replacing with the L1-norm will cause some classifier weights to shrink to zero, thus doing feature selection for the classifier.

Q12 ANSWER:

- [4 marks] The L1 loss will make the training robust to outliers that don't follow the linear trend.
- [6 marks] Those points with large errors will be tolerated due to the linear trend of the loss function – i.e., large errors are treated equally as small errors, unlike ordinary least squares, which squares the error, and thus large errors have more influence.

Q13 ANSWER:

- [3 marks] The minimum at $z=1$ will try to make the classifier predict $f(x)=+1$ or $f(x)=-1$ only.
- [2 marks] linear loss when $z<1$, means it will be robust to misclassified outliers.
- [3 marks] Increasing loss when $z>1$, means it will also try to push well-classified examples back to $f(x)=1$. Sensitive to "easy" or "too correct" examples.
- [2 marks] This will not be a good classifier, since it tries to push all easy examples to $z=1$.

--- END ---