FIT 3181/5215 Deep Learning

Quiz for:

Introduction to Machine Learning

Tutor Team

Department of Data Science and AI
Faculty of Information Technology, Monash University

Content

The quizzes cover the material in Lecture 01 and Tutorial 02 which include:

- Machine learning pipeline
- Evaluation metric: precision, recall, F1-score

Question 1

Consider the problem of predicting the house price in Melbourne. Which of the following statements are most likely true (MC).

- A. This problem should be approached from a supervised learning problem
- B. This problem is better to be addressed as a classification problem
- C. This problem is better to be addressed as a regression problem
- D. We should adopt abnormality detection methods for this prediction problem

Question 2

John class has 50 students. He wants to build a machine learning model which helps to recognize students' face and automatically records their attendance in his class. He lists some must-do tasks as below. Choose the correct pipeline? (SC)

(1) *Taking students' pictures and upload to internet*	(2) *Building a machine learning model*
(3) *Grouping student's pictures and their ID number (e.g. student ID 01: image A, B; student ID 02: image C, D, etc)*	(4) *Collecting student's ID number*
(5) *Splitting dataset into training set, testing set*	(6) *Download face images from internet*
(7) *Filtering, removing low quality images*	(8) *Training model on training set;*
(9) *Testing model on testing set*	(10) *Deploying model into a real hardware system*

- A. (1) (6) (3) (4) (7) (5) (2) (8) (9) (10)
- B. (1) (6) (7) (4) (3) (2) (5) (8) (10) (9)
- C. (1) (6) (7) (4) (3) (5) (2) (8) (9) (10)
- D. (1) (6) (3) (4) (7) (5) (2) (8) (10) (9)

3 cant come before 4

Question 3

John's class has 50 students. He wants to build a machine learning model which helps to recognize students' face and automatically records their attendance in his class. John splits his students' images into two sets: training set and testing set. John ***trains*** *three models A, B, C on* ***training set*** *and choose the model with the* ***highest accuracy*** *on* ***training set*** *as the final model. Which are correct statements? (SC)*

- A. Join's model will be possibly experiencing an overfitting problem
- B. Join's model will be possibly experiencing a dataset biasing problem
- C. The dataset was used wrongly
- D. None of above

Question 4

John wants to build a machine learning model which helps to recognize a strange student attending his class? The confusion table is as below. Choose the correct statement? (MC)

True labels

Predicted as	Stranger	John's student
Stranger	2	5
John's student	3	40

- A. John's class has total 50 students
- B. John's model accuracy (2+40) / (2+3+5+40) = 84%
- C. John model's recall 2 / 5 = 40%
- D. John model's precision 2/7

strange student is not John's student

2/(2+3)

Question 5

John wants to build 2 models A, B which helps to recognize a strange student attending his class? The confusion as the tables below. Choose the correct statement? (SC)

Model A

Predicted as	True labels: Stranger	True labels: John's student
Stranger	4	7
John's student	1	88

Model B

Predicted as	True labels: Stranger	True labels: John's student
Stranger	2	3
John's student	3	92

- A. Model A is better than model B in prediction accuracy
- B. Model A is better than model B in recall
- C. Model A is better than model B in true negative rate
- D. Model A is better than model B in precision

Question 6

- *Our dataset has 10 spam emails (label 1) and 50 non-spam emails (label -1). Given an email x, the classifier returns $p(x) = P(y = 1|x)$ as the probability to assign x to the class 1. The classifier assigns x to the class 1 if $p(x) \geq 0$. Choose the correct statements.* (MC)

- A. Recall = 1 10 / (10+0)
- B. Precision = 1/6 10/(10+50)
- C. True Positive Rate = 0 10/P = 10 / (10+0) = 1
- D. False Positive Rate = 1

fp/(fp+tn) = 50 / (50+0)

- Consider **a threshold** $\theta \in [0; 1]$. Given a data point x, the ML model returns $p(x) = \mathbb{P}(y = 1|x)$ (i.e., the **probability to predict** x as the label 1). We predict x as the **label 1** if $p(x) = \mathbb{P}(y = 1|x) \geq \theta$ and the **label -1** otherwise.

10	50
0	0

Question 7

- *Our dataset has 10 spam emails (label 1) and 50 non-spam emails (label -1). Given an email x, the classifier returns $p(x) = P(y = 1|x)$ as the probability to assign x to the class 1. The classifier assigns x to the class -1 if $p(x) \leq 1$. Choose the correct statements.* (MC)

- A. Recall = 0 0/(0+10)
- B. Accuracy = 5/6
- C. True Positive Rate = 0 same as recall
- D. False Positive Rate = 1

0	0
10	50

Question 8

Referring to the segment of codes below, what is the number of data points return to variable train_X if there are 200 data points in X? (SC)

```
from sklearn.model_selection import train_test_split
train_X, test_X, train_y, test_y = train_test_split(X, y, test_size = 0.3, random_state=2)
```

- A.60
- B. 70
- C. 140
- D. 200

Question 9

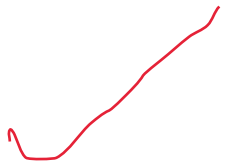
Choose the possible solution If a model performs great on the training data but generalizes poorly to new instance. (MC)

- A. Selecting a more complicated algorithm, increasing the number of parameters
- B. Reducing the number of input features
- C. Increasing the amount of training data
- D. Increasing testing set and reducing training set

Question 10

- *Assume that we have 4 classes in {cat = 1,dog = 2,lion = 3, monkey = 4}. What is one-hot label of categorical label "**lion**"?*

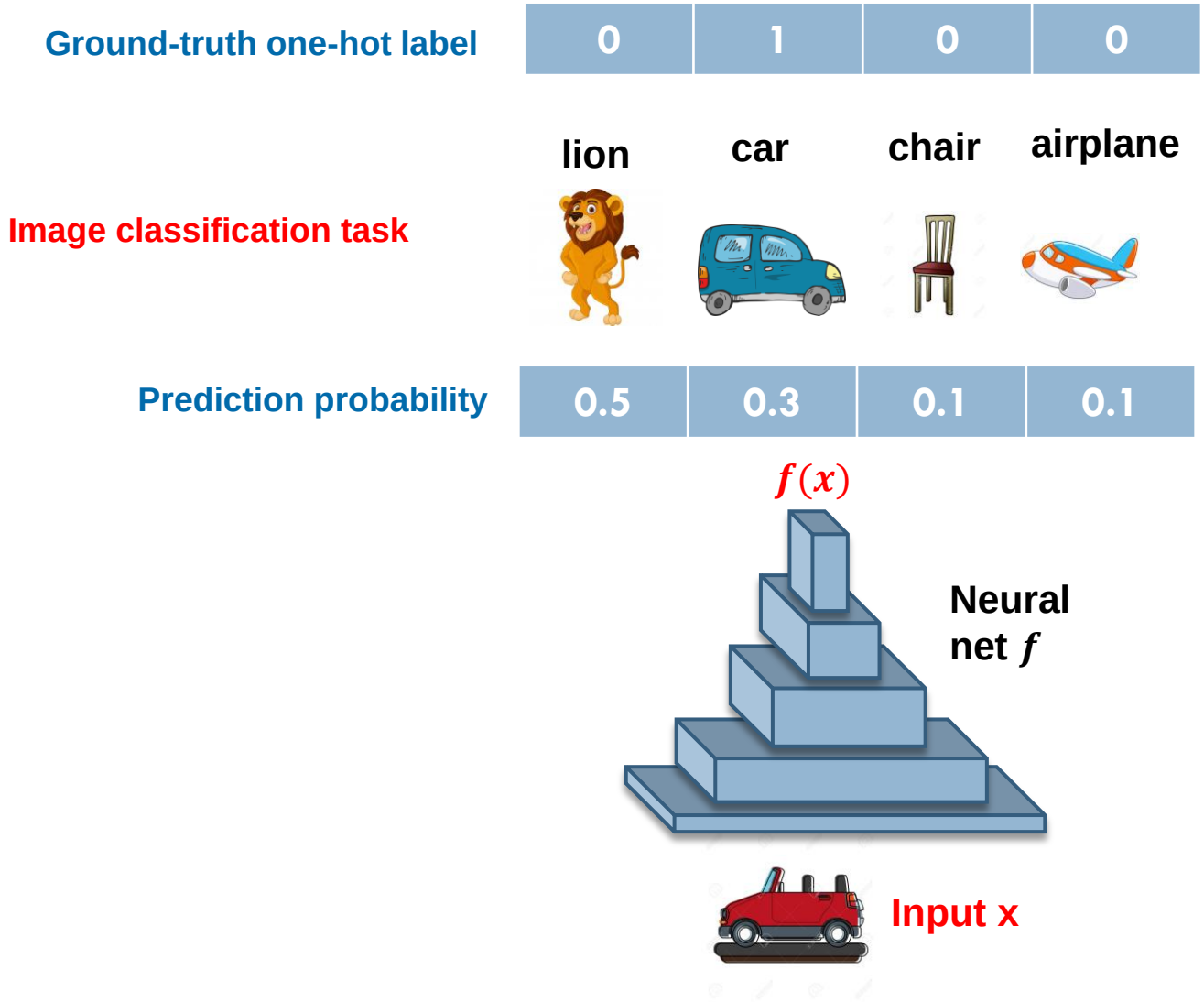
- A. [1,0,0,0]
- B. [0,1,0,0]
- C. [0,0,1,0]
- D. [0,0,0,1]



Question 11

- *Consider an image classification task as showing. Assume that the model give a* ***prediction probabilities*** $f(x) = [0.5, 0.3, 0.1, 0.1]$ *and* ***categorial ground-truth label*** *of* x *is car. What is the cross-entropy loss suffered by this prediction?*

- A. $-\log 0.5$
- B. $\log 0.3$
- C. $-\log 0.3$
- D. $\log 0.5 + \log 0.1$



Question 12

- *Assume that we have **4 classes** in **{cat = 1,dog = 2,lion = 3, monkey = 4}**. Given a data example x with **ground-truth label "dog"**, assume that a machine learning model gives **discriminative scores** to this x as $h_1 = -3$, $h_2 = 10$, $h_3 = 5$, $h_4 = 0$. Choose all correct answers. (MC)*

- A. The model predicts x as lion
- B. The model predicts x as dog
- C. This is a correct prediction
- D. This is an incorrect prediction

Question 13

- *Assume that we have **4 classes** in **{cat = 1,dog = 2,lion = 3, monkey = 4}**. Given a data example x with **ground-truth label "dog"**, assume that a machine learning model gives **discriminative scores** to this x as $h_1 = -3$, $h_2 = 10$, $h_3 = 5$, $h_4 = 0$. What is the probability to predict x as lion or $p(y = lion \mid x)$?*

- A. $\frac{e^5}{e^{-3}+e^{10}+e^5+e^0}$
- B. 1
- C. $\frac{e^0}{e^{-3}+e^{10}+e^5+e^0}$
- D. $\frac{e^{10}}{e^{-3}+e^{10}+e^5+e^0}$

Question 14

- *Assume that we have **4 classes** in **{cat = 1,dog = 2,lion = 3, monkey = 4}**. Given a data example x with **ground-truth label "cat"**, assume that a machine learning model gives **discriminative scores** to this x as $h_1 = -3$, $h_2 = 10$, $h_3 = 5$, $h_4 = 0$. What is the CE loss suffered by this prediction?*

- A. $\log \frac{e^5}{e^{-3}+e^{10}+e^5+e^0}$
- B. $\log \frac{e^{-3}}{e^{-3}+e^{10}+e^5+e^0}$
- C. $-\log \frac{e^{-3}}{e^{-3}+e^{10}+e^5+e^0}$
- D. $-\log \frac{e^{10}}{e^{-3}+e^{10}+e^5+e^0}$