**1. Which algorithms belong to discrete supervised learning?** \*

K-means clustering

Logistic Regression

Perceptron

SVM

PCA

**2. Decision trees belong to which tribe of machine learning?\***

Symbolists

Evolutionaries

Bayesians

Connectionists

Analogizers

**3. There are 10 classes in MNIST dataset (0 ~ 9). What is the output dimension of last layer if we use SoftMax as activation function? \***

2

10

No need for activation function

Any number

**4. Can perceptron be used for learning XOR operation? \***

Yes

No

**5. What is the meaning of LLM in AI?**

Master of Law

Law of Large Numbers

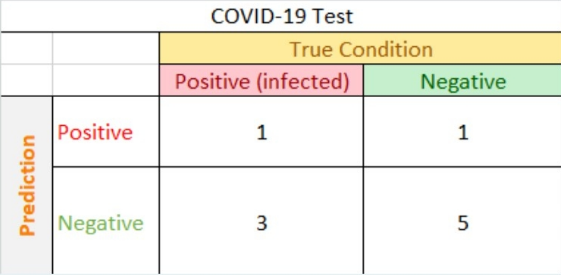
Large Llama

Large Language Models

**6. How many SVM classifiers are required for a 10-class problem if we use one-against-one strategy?** 45

**7. How many SVM classifiers are required for a 1000-class problem if we use one-against-all strategy?** 1000

**1.Given the confusion matrix of a COVID-19 test, what is its precision?**  0.5



**2.What is the best metric to evaluate model performance using Precision-Recall curve?**

Precision + Recall

Precision - Recall

Area under PR curve

Precision \* Recall

**3.Prediction Error can be decomposed into**

Bias

Variance

Irreducible error

training error

**4.How to reduce Overfitting?**

Regularization

K-Fold cross validation

K-means clustering

K Nearest Neighbor

KFC

Reduce model size

**5.Which statement about L1 and L2 norm for regularization is FALSE?**

Both L1 norm and L2 can be used for regularization

L1 and L2 norm tend to make training accuracy lower

L2 is preferred to be used for feature selection

Linear regression + L1 norm regularization = LASSO

**1. Two vectors A and B are orthogonal if**

A and B are both zero

Dot product of A and B is equal to 0

Product of A and B are invertible

Dot product of A and B is equal to an identify matrix

**2. Suppose the probability of rainy weather is 0.2 and sunny weather is 0.8. The probability that a student goes to school on a rainy day is 0.12, and goes to school on a sunny day is 0.48. What are the chances that students will decide to stay home regardless of the weather?**

0.1

0.2

0.4

0.6

**3. If event A and B are mutually exclusive. What is the conditional probability P(A|B)?**

P(A)

P(AB)

P(B)

0

**4. The probabilty of event A P(A) = 0.4, event B is P(B) = 0.8. If event A and B are independent. What is the probability of both A and B happen?**

0.32

**5. What is the expected value of a fair coin?**

0.5

**6. Given a square matrix A, suppose Ax=λx, A is a scalar and x is a non-zero vector. Which statements below are TRUE?**

x is an eigenvector of A

λ is an eigenvalue of A

λ is the eigenvalue corresponding to eigenvector A

Matrix A does not change the direction of vector x

**7. Which statements below are TRUE?**

All square matrices are invertible.

The dot product of two vectors is a scalr

The outer product of two vectors is a matrix

PCA assumes that the principal components are orthogonal

**期中模擬考偶數**

**1.Which of the following do you typically see as you move to deeper layers in a ConvNet?**

nH and nW decrease, while nC decreases

nH and nW increase, while nC decreases

nH and nW decrease, while nC increases

nH and nW increase, while nC increases

**2.The concept "momentum" in gradient descent optimization can be used to**

Find the global minimum

Avoid getting stuck in local minimum

Find all local minimums

Convert a function into a convex function

**3.Suppose your input is a 300 by 300 color (RGB) image, and you use a convolutional layer with 100 filters that are each 3x3. How many parameters does this hidden layer have (including the bias parameters)?**

2800

**4.Suppose your input is a 200 by 200 color grayscale image, and you have first fully connected hidden layer with 100 neurons. How many parameters does this hidden layer have (including the bias parameters)?**

4000100

**5.Which statements on Residual Networks are true?**

Using a skip-connection helps the gradient to backpropagate and thus helps you to train deeper networks

A ResNet with L layers would have L^2 skip connections in total.

The skip-connections compute a complex non-linear function of the input to pass to a deeper layer in the network.

The skip-connection makes it easy for the network to learn an identity mapping between the input and the output within the ResNet block.

**6.Training error can be viewed as a bias-variance tradeoff. In general, which type of error is larger when the model complexity decreases (underfitting)?**

Bias

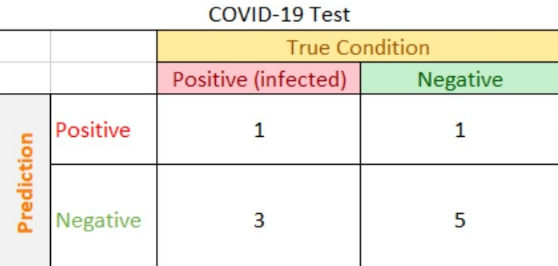
Variance

Both

None

**7.Here is a COVID-19 test result. The rows are predictions while the columns are true condition.**

**Positive means the patient is infected. What is the Sensitivity? (Please enter number between 0 and 1)**



0.25

**Which algorithms are unsupervised algorithms**

k-means clustering

word2vec (Word Embedding)

LSTM

Spectrum clustering

正確10/10點數

**What is the value of z in the Python code below?**

import numpy as np

w = np.array([[1,1,0], [0, 1,1]])

x = np.array([[1,0],[0,1],.[1,0]])

y = np.matmul(w, x)

z = np.sum(y)

4

**10.You are training a classifier on a highly imbalanced dataset. The majority class accounts for 99% of the training data, and your classifier achieved 99% accuracy. Which statements are true?**

Accuracy metric is not a good idea for imbalanced class problems

Accuracy metric is a good idea for imbalanced class problems

Precision and recall metrics are good for imbalanced class problems

Precision and recall metrics aren't good for imbalanced class problems

**11.Given a training sample, suppose its true label y=1, and our prediction p = 0.7, what is the binary cross entropy loss?**

-0.3

-0.7

-log(0.3)

-log(0.7)

-0.7\*log(0.7)

**12.Which type of neural networks can be used for time-series data?**

RNN

LSTM

1D CNN

GRU

**13.Two events X and Y are independent if**

P(XY) = 0

P(XY) = P(X)P(Y)

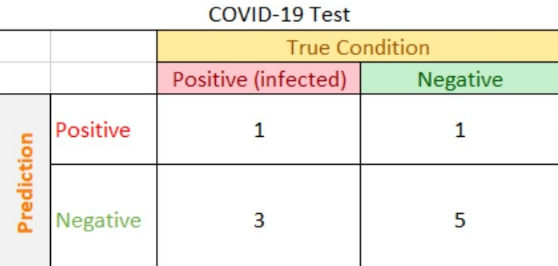
P(XY) = P(X)

P(XY) = P(Y)

**期中模擬考奇數**

**1.Here is a COVID-19 test result. The rows are predictions while the columns are true condition.**

**Positive means the patient is infected. What is the Sensitivity? (Please enter number between 0 and 1)**



0.25

**2.You are training a classifier on a highly imbalanced dataset. The majority class accounts for 99% of the training data, and your classifier achieved 99% accuracy. Which statements are true?**

Accuracy metric is not a good idea for imbalanced class problems

Accuracy metric is a good idea for imbalanced class problems

Precision and recall metrics are good for imbalanced class problems

Precision and recall metrics aren't good for imbalanced class problems

**3.Suppose your input is a 300 by 300 color (RGB) image, and you use a convolutional layer with 100 filters that are each 3x3. How many parameters does this hidden layer have (including the bias parameters)?**

2800

**4.Which algorithms are unsupervised algorithms**

k-means clustering

word2vec (Word Embedding)

LSTM

Spectrum clustering

**5.Two events X and Y are independent if 4**

P(XY) = 0

P(XY) = P(X)P(Y)

P(XY) = P(X)

P(XY) = P(Y)

**6.The concept "momentum" in gradient descent optimization can be used to 대**

Find the global minimum

Avoid getting stuck in local minimum

Find all local minimums

Convert a function into convex function

**7. What is the value of z in the Python code below?**

import numpy as np

w = np.array([[1,1,0], [0, 1, 1]])

x = np.array([[1,0].[0,1],.[1,0]])

y = np.matmul(w, x)

z = np.sum(y)

4

**8.** **Which type of neural networks can be used for time-series data?**

RNN

LSTM

1D CNN

GRU

**9. Training error can be viewed as a bias-variance tradeoff. In general, which type of error is larger when the model complexity decreases (underfitting)?**

Bias

Variance

Both

None

**10.** **Suppose your input is a 200 by 200 color grayscale image, and you have first fully connected hidden layer with 100 neurons. How many parameters does this hidden layer have (including**

**the bias parameters)?**

4,000,100

**11. Which of the following do you typically see as you move to deeper layers in a ConvNet?**

nH and nW decrease, while nC decreases

nH and nW increase, while nC decreases

nH and nW decrease, while nC increases

nH and nW increase, while nC increases

**12.** **Which statements on Residual Networks are true?**

Using a skip-connection helps the gradient to backpropagate and thus helps you to train deeper networks.

A ResNet with L layers would have L^2 skip connections in total.

The skip-connections compute a complex non-linear function of the input to pass to a deeper layer in the network.

The skip-connection makes it easy for the network to learn an identity mapping between the input and the output within the ResNet block.

**13.Given a training sample, suppose its true label y=1, and our prediction p = 0.7, what is the binary cross entropy loss?**

-0.3

-0.7

-log(0.3)

-log(0.7)

-0.7\*log(0.7)

\*