机器学习复习题

一、判断题

1. ID3 algorithm can successfully avoid overfitting in any case.（　　）
2. C4.5 algorithm tends to split a node with an attribute that maximizes information gain ratio.（　　）
3. The *K*-means algorithm belongs to supervised learning.（　　）
4. *K*-NN is a clustering algorithm.（　　）
5. *K*-means algorithm can converge to global optimum.（　　）

二、选择题

1. The main reasons of pruning in decision trees **NOT** include（　　）

A. Reduce tree size

B. Improving training accuracy

C. Avoid overfitting

D. Improving test accuracy

2. Which of the following items is **FALSE** about decision trees? (　　)

A. One way of controlling overfitting is to regularize the hypothesis selection.

B. In any case, C4.5 algorithm algorithm can avoid overfitting.

C. One way of controlling overfitting is to limit the max depth of trees.

D. C4.5 algorithm tends to split a node with an attribute that maximize information gain ratio.

3. Which of the following statements correctly describe *K*-means?（　　）

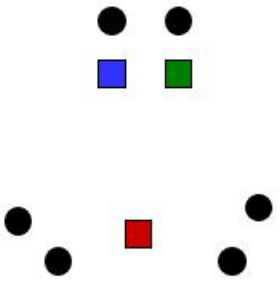
A. Total square distance to class center keeps decreasing

B. Converge to global optimum

C. Optimal number of clusters *K* easy to find

D. Need to run multiple times with random class center

4. For the initialization of *K*-means shown in the following graph, what sample distribution will we get when it converges?（　　）



A. 2 samples in each class

B. All 6 go to one single class

C. 1 each in blue and green, other 4 in red

D. It can not be predicted

三、解答题

1. The table below lists the *X* and Y coordinatesof cities A to E.

a. Calculate the Manhattan distances between the cities.

b. Use the single link agglomerative clustering algorithm to calculate the merging sequence of cities A-E and draw the dendrogram.

c. Now we do not consider city B. In the remaining cities, take $K = 2$ and cities A and E as the initial cluster centers, describe the clustering steps by the K-means algorithm. (Note in *K*-means, we use the Euclidean distance).

	X	*Y*
A	10	5
B	20	20
C	30	10
D	30	15
E	5	10

2. The distances between 5 cities are listed below.

	A	B	C	D	E
A	0	200	160	190	300
B		0	80	170	290
C			0	240	210
D				0	120
E					0

a. Using agglomerative clustering method with the single link, list the merging sequence and draw the dendrogram.

If City A is classified as type I, and Cities B and C are classified as type II, then according to 1-NN (take $K = 1$ in *K*-NN), what types should City D and City E be classified, respectively? What type should City E be classified according to 2-NN? Write the steps.

3. The table below lists the *X* and *Y* coordinates of cities *A* to *H*, we take *K*-means algorithm with $K = 3$ using Euclidean distance, and select the cities *A*, *D* and *G* as the initial cluster centers, calculate three clusters and the corresponding coordinates of new centers for the first clustering.

	X	*Y*
A	2	10
B	2	5
C	8	4
D	5	8

E	7	5
F	6	4
G	1	2
H	4	9

4. In the table below, there is a classification problem with target “play tennis”, please calculate the entropy for the classification that is whether or not to play tennis. In ID3 algorithm, which is the first attribute to split on. (take logarithm with base 2: $\log_2$, $\log_2 3=1.5850$).

Day	Outlook	Temp.	Humidity	Wind	Play?
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
3	Overcast	Hot	High	Weak	Yes
4	Rain	Mild	High	Weak	Yes

搜索复习题

一、判断题

1. Regarding the concept of mini-max rule for a zero-sum game, each player maximizes the minimal loss possible for the other.（　　）

2. Chinese chess is a zero-sum game.（　　）

3. General games are only adversarial and competitive.（　　）

二、选择题

1. In the following items, which is **NOT** a deterministic, zero-sum game?（　　）

A. Tic-tac-toe

B. Pancake problem

C. Chinese chess

D. Checkers

2. Which of the following is a **TRUE** statement about mini-max algorithm?（　　）

A. The mini-max algorithm performs a greedy search of the game tree.

B. The mini-max algorithm performs a local search of the game tree.

C. The mini-max algorithm performs a complete depth-first exploration of the game tree.

D. The mini-max algorithm performs a complete breadth-first exploration of the game tree.

3. Which of the following items is **FALSE** about zero-sum games? ( )

A. Pure competition

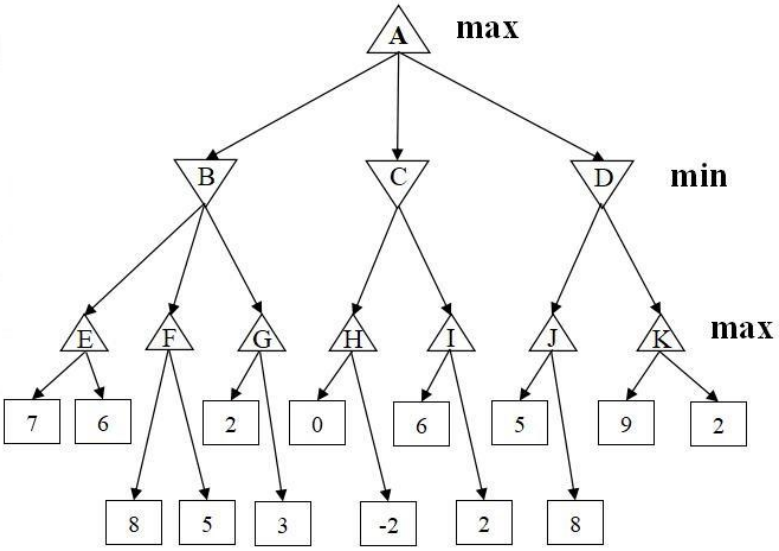
B. Adversarial

C. Cooperation and indifference

D. Have opposite utilities

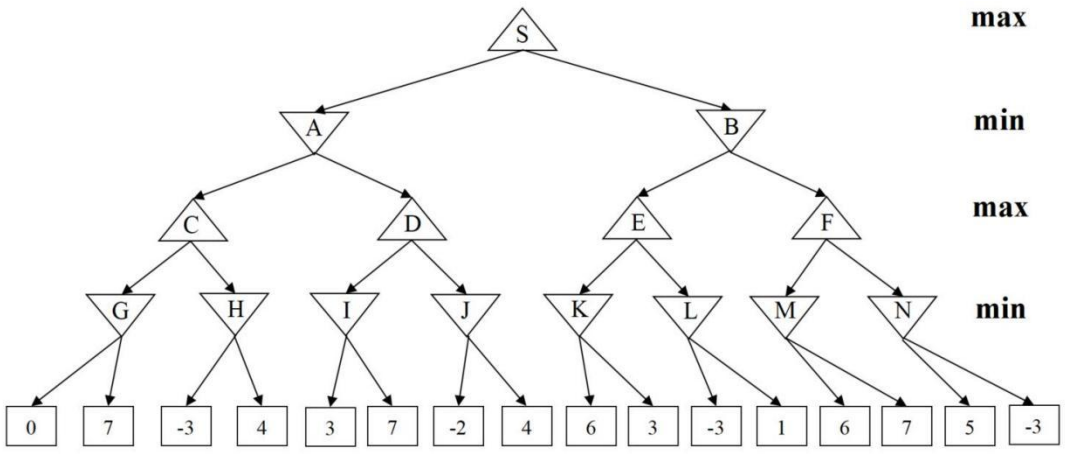
三、填空题

1. Fill in the missing values for the mini-max search tree: Top cell (max)= _____(A), Middle cells in the second level (min)=_____(B), _____(C), _____(D) (from left to right), Middle cells in the third level (max)=_____(E), _____(F), _____(G), _____(H), _____(I), _____(J), _____(K) (from left to right).

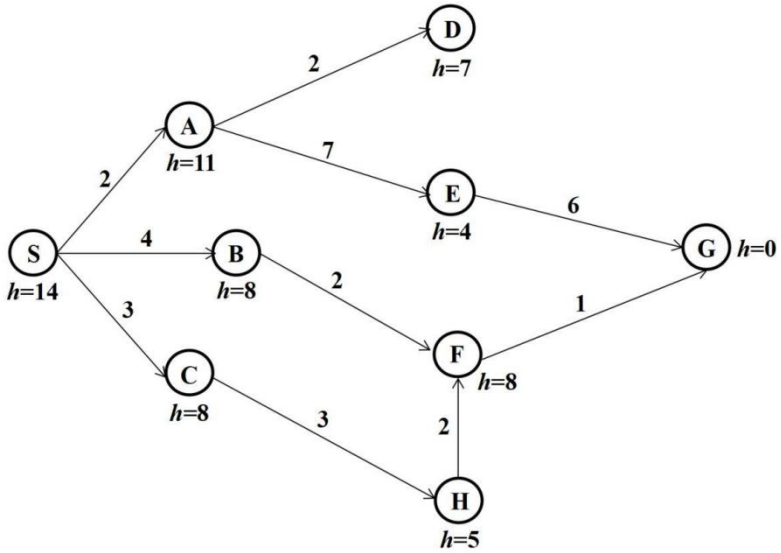


四、解答题

1. Calculate the missing values S, A, C, F, I, M for the mini-max search tree, and give the leaf nodes which are not necessary to be visited, according to α-β (alpha-beta) pruning algorithm.



2. The figure below shows a sear problem, where *S* is the start state and *G* is the end state. The number next to each node is the heuristic value of that node, and the number next to each edge is the cost of the edge. Describe the search steps from *S* to *G* using *A** search, and is the heuristic function given in this problem admissible?



3. Given a 3-puzzle problem, the start state and goal state are presented below, we take *A** search to solve this problem. In *A** search, the cost function $g(n)$ is defined by the number of moves of the empty tile (without any number) , the heuristic function $h(n)$ is defined by the total Manhattan distance, the moving operators are defined as *N* (↑)、*E* (→)、*W* (←) and *S* (↓), please give the search tree for *A** search and the function values $f(n)$ for each step.

3	1
2	

Start State

1	2
3	

Goal State

神经网络复习题

一、判断题

1. Traditional computer vision methods have clear advantages over neural networks in the ImageNet image classification task.（　　）

2. One of the advantages of CNN over the fully-connected neural network is its smaller parameter size.（　　）

3. Activation functions in a neural network are usually linear because of the simpler form.（　　）
4. A linear functions can be used as an activation function, since a large enough neural network can be used to approximate any function, including nonlinear ones.（　　）

二、选择题

1. What is an advantage of the fully convolutional network？（　　）
A. It outputs a label for the input image
B. It makes predictions for all pixels at once, in comparison to the sliding window method
C. It uses only convolution and pooling operations, just like CNN
D. It works as well as CNN for image classification
2. The popularity of CNN in image processing is **NOT** based on that（　　）
A. It can detect small patterns instead of the whole image
B. It uses shared parameters when processing different regions
C. It can handle the same object in various sizes
D. It can remember the past in sequential input
3. Which of the following statement is TRUE for CNN in image classification？（　　）
A. Each unit processes the whole input image all at once.
B. Max pooling does not change the size of its input.
C. Parameters of a convolution kernel are the same for all different regions.
D. Its last layer does not use the Softmax operator.
4. Which statement is **NOT** correct for CNN（　　）
A. It can detect small patterns instead of the whole image
B. It uses shared parameters when processing different regions
C. It can handle the same object in various sizes
D. It is not a good choice when processing text
5. Which of the following items is **TRUE** about neural networks? (　　)
A. Sigmoid function is an appropriate activation function.
B. Given any discontinuous function $f(x)$, if a 2-layer neural network has enough hidden units, then there is a choice of weights that allow it to closely approximate $f(x)$.
C. Linear function is an appropriate activation function.
D. A two-layer neural network with a sufficient number of neurons can approximate any function to any desired accuracy.
6. For the confusion matrix below, $F1$ value of retrieval is (　　).

Actual\Prediction	**Selected**	**Not selected**
Relevant	800	3200
Irrelevant	200	1800

（A）0.22　　（B）0.28　　（C）0.32　　（D）0.36

三、解答题

1. The input is a 4×4 matrix, it first goes through a 2×2 convolution kernel with zero padding and stride 1, then a 2×2 max pooling is applied with stride 1. The output goes through a one-layer neural network with softmax, and final output is the probability for two classes.

0	1	2	0
1	3	0	1
0	2	4	1
0	0	2	1

Input matrix

1	2
-1	0

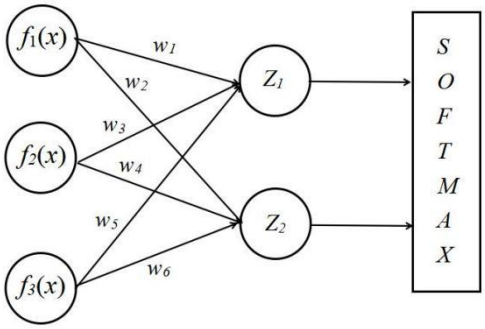
Convolution kernel

(1) Calculate the outputs from the convolution step and after the max pooling step.

(2) Flatten the output into a one-dimensional vector, with row major order, then multiply it by the weight matrix below. Use sigmoid function to convert the output to probability.

0.1	0.2
0.2	0.3
-0.3	0.1
0.5	-0.4

2.In the following neural network, set $[f_1(x), f_2(x), f_3(x)] = [1, 0, 2]$, $[w_1, w_2, w_3, w_4, w_5, w_6] = [0.2, 0.8, 1.6, 1.8, -0.1, -0.9]$, and let the activation function for z_1 and z_2 be the sigmoid function, calculate the final softmax result.



决策网络复习题

一、判断题

1. Value of Perfect Information (VPI) does **NOT** have the following property:（　　）

A. Non-negative　　B. Non-additive

C. Non-commutative　　D. Order-independence

2. Which of the following can be represented as a chance node?（　　）

A. To be or not to be

B. Be an early bird or a night owl

C. Tomorrow is rainy or sunny

D. Play the dice game or not

二、解答题

1. A company owns a ship worth 2000W¥. The ship will depreciate in value by 10W¥ during the coming year. The company has been offered an accident insurance policy for 20W¥ per year that the company will be given 1600W¥ if it has an accident. The company assesses the risk of the ship having an accident once during the year as 1% and the risk of two or more accidents to be negligible.

Suppose an accident means the ship's value decreases to 0. Calculate the expected losses the company insure or not insure in the coming year. Should the company insure?. If the accident risk increases to 2%, should the company insure?

2. A student owns a new bicycle worth ¥200 which she uses every day. Owing to wear and tear the bicycle will depreciate in value by ¥50 during the coming year. She has been offered a "new for old" insurance policy for ¥30 per year. Given that the risk of two or more thefts can be negligible and that the risk of the bicycle being stolen once during the year is 5%, determine the best action (insure or not) she could perform.