浙江大学 2016 –2017 学年 春夏 学期

《Artificial Intelligence》课程期末考试试卷

课程号： 21191890 ，开课学院： 计算机科学与技术学院

考试试卷：**A 卷**、~~B 卷~~（请在选定项上打√）

考试形式：**闭**、~~开~~卷（请在选定项上打√），允许带_____入场

考试日期： 2017 年 6 月 25 日,考试时间： 120 分钟

诚信考试，沉着应考，杜绝违纪。

考生姓名：_____学号：_____所属院系：_____

题序	一	二	三	四	五	六	七	八	总 分
得分									
评卷人									

1. Fill in the blanks (30 points)

1) Two common structures are used, namely first-in-first-out (FIFO queue) and last-in-first-out (LIFO queue). The breadth-first-search uses a _____queue, depth-first search uses a_____ queue.

2) In alpha-beta pruning search, the algorithm maintains two values, alpha and beta, which represents the maximum score that the maximizing player is assured of and the minimum score that the minimizing player is assured of respectively. At the beginning of alpha-beta search, alpha is set to _____ and beta is set to _____, i.e. both players start with their lowest possible score.

3) A and B are two random variables. P(A) and P(B) are their probabilities. It is known that $P(A) + P(B) = 0.5$, and $\frac{P(A|B)}{P(B|A)} = \frac{1}{4}$. Then the $P(A)$ is_____.

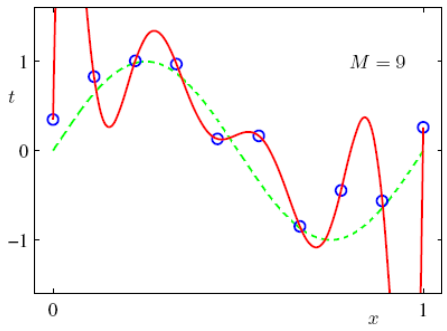
4) In your 10-day vacation in Alaska, you kept the following log on the weather and whether you saw a bear that day:

(rain, bear)	1 day
(¬rain, bear)	2 days
(rain, ¬bear)	6 days
(¬rain, ¬bear)	1 day

a) Compute the marginal probability $P(bear)$ = _____

b) Compute the conditional probability $P(\neg bear|rain)$ = _____

5) In the figure, the circles show a plot of a training data set of 10 data points, the dash line shows the function f(x) used to generate the data and solid curve shows the higher order polynomial g(x) fitted to given 10 data points. The fitted curve passes exactly through each data point, so the value of RMS error E_{RMS} = 0 and g(x) gives a very poor representation of the function f(x). This behavior is known as _____ (please select over-fitting or under-fitting to fill in this blank).



6) For a given likelihood function $p(x_n|\theta)$, if we obtain a data set of observations $X = \{x_1, x_2, x_3\}$ and these data points are independent and identically distributed (i.i.d.), then $p(X|\theta) = p(x_1, x_2, x_3|\theta)$= _____.

7) For multivariate Gaussian distribution N(x|μ, Σ) of the D dimensional input space x, we have _____ independent parameters for μ and Σ. If Σ is a diagonal matrix and $\Sigma = \sigma^2\mathbf{I}$, the number of total parameters reduces to _____.

8) The Linear basis function models involve linear combinations of fixed nonlinear functions of the input variables. If given basis functions $\boldsymbol{\phi}(\mathbf{x}) = (\phi_0(\mathbf{x}), \phi_1(\mathbf{x}), \phi_2(\mathbf{x}))^T$, where $\phi_0(\mathbf{x}) = 1$ and the model parameters $\mathbf{w} = (w_0, w_1, w_2)^T$, then the linear basis function $y(\mathbf{x}, \mathbf{w})$ = _____.

9) In general, a deep convolutional neural network consists of convolutional layer, pooling layer, fully-connected layer and classifier layer, the softmax is usually employed at the _____ layer.

10) Reinforcement learning mainly consists of policy, value function and model. A_____ maps a state to an action, and a value function is a prediction of future reward. In Q-value function, discount factor γ is usually used, the range of discount factor γ is _____.

2. Multiple Choice (36 points, only one of the options is correct.)

1) Consider three 2D points a = (0, 0), b = (0, 1), c = (1, 0). Run k-means with two clusters. Let the initial cluster centers be (-1, 0), (0, 2). What clusters will k-means learn after one iteration? _____

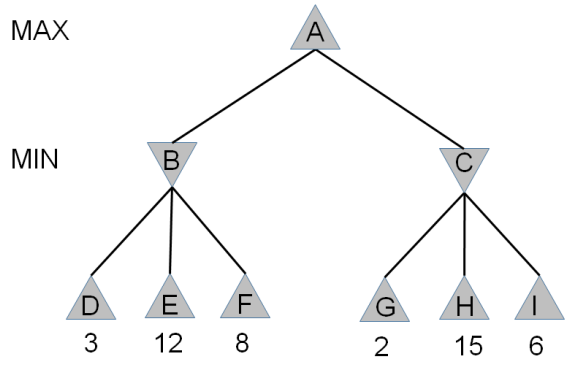
(A) $\{a\}, \{b, c\}$
(B) $\{a, b\}, \{c\}$
(C) $\{a, c\}, \{b\}$
(D) none of the above

2) The sigmoid function in a neural network is defined as $g(x) = \frac{e^x}{1+e^x}$. There is another commonly used activation function called the hyperbolic tangent function, which is defined as $\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$. How are these two functions related?

_____

(A) $tanh(x) = g(x) - 1$
(B) $tanh(x) = 2g(x) - 1$
(C) $tanh(x) = g(2x) - 1$
(D) $tanh(x) = 2g(2x) - 1$

3) Which nodes will be pruned along with their branches by alpha-beta pruning? _____



(A) I (B) H, I (C) G, H, I (D) C, H, I

4) Consider a 3-puzzle where, like in the usual 8-puzzle game, a tile can only move to an adjacent empty space. Given the initial state which of the following state **cannot** be reached? _____

(A) 

(B) 

(C) 

(D) 

5) Given two Gaussian distribution $N(x|-1,1)$ and $N(x|1,1)$, which of the following formula is correct? _____

(A) $N(0|-1,1) > N(0|1,1)$

(B) $N(-1|-1,1) > N(-1|1,1)$

(C) $N(0|-1,1) < N(0|1,1)$

(D) $N(-1|-1,1) < N(-1|1,1)$

6) The Fisher's criterion is defined to be _____

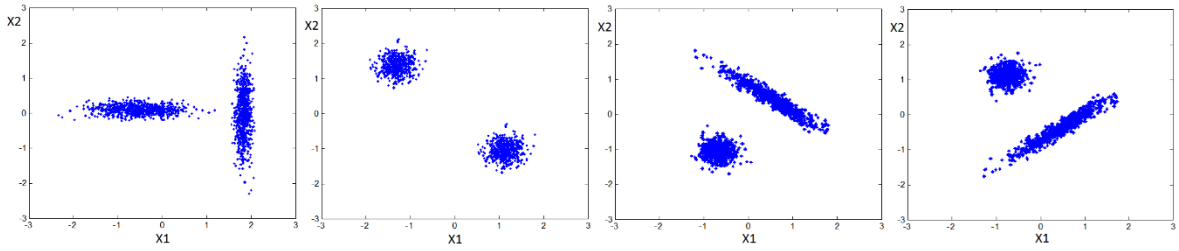
(A) the separation of the projected class means.

(B) the separation of the projected class variances.

(C) the ratio of the between-class variance to the within-class variance.

(D) the ratio of the within-class variance to the between-class variance.

7) Suppose we have a data set $\{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ drawn from the mixture of two 2D Gaussians, which can be written as $p(\mathbf{x}) = 0.5N(\mathbf{x}|\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1) + 0.5N(\mathbf{x}|\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$. If $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \sigma^2\mathbf{I}$ in this model, which of the following figures is consistent with the distribution of data points $p(\mathbf{x})$? _____



(A) (B) (C) (D)

8) Consider a polynomial curve fitting problem. If the fitted curve oscillates wildly through each point and achieve bad generalization by making accurate predictions for new data, we say this behavior is over-fitting. Which of the following methods **cannot** be used to control over-fitting? _____

(A) Use fewer training data

(B) Add validation set, use Cross-validation

(C) Add a regularization term to an error function

(D) Use Bayesian approach with suitable prior

9) AlexNet (one of popular multi-layer convolutional neural networks for image classification) is trained in a_____ setting, K-means clustering is employed in a _____ setting and Boosting for classification is implemented in a_____ setting , linear regression model for classification is realized in a _____ setting.

(A) unsupervised, supervised, supervised, unsupervised

(B) supervised, supervised, supervised, supervised

(C) supervised, supervised, unsupervised, unsupervised

(D) supervised, unsupervised, supervised, supervised

10) In linear least square regression model, we can add a regularization term to an error function (i.e., sum-of-squares) in order to control _____. The lasso regularizer will introduce_____ solution compared to quadratic regularizer.

(A) over-fitting, dense (B) over-fitting, sparse

(C) under-fitting, dense (D) under-fitting, sparse

11) We can decompose the expected squared loss of one predict model as follows:

$$\text{expected loss} = (\text{bias})^2 + \text{variance} + \text{noise}$$

In general, one flexible model (i.e., under-fitting) model will introduce high_____ and one rigid model (i.e., overfitting) model will introduce high _____. There is a tradeoff between a model's ability to minimize bias and variance.

(A) variance, bias (B) bias, bias

(C) variance, variance (D) bias, variance

12) Which description is **not** correct in terms of supervised learning, semi-supervised learning, unsupervised learning and reinforcement learning?_____

(A) Reinforcement learning is one of specific supervised learning methods.

(B) Semi-supervised learning falls between unsupervised learning (without any labeled training data) and supervised learning (with completely labeled training data).

(C) Reinforcement learning is neither supervised learning nor unsupervised.

(D) Deep reinforcement learning is a combination of deep learning and reinforcement learning.

13) In reinforcement learning, Q-learning defines a function Q(s, a) representing the _____ when we perform action ***a*** in state ***s***, and continue optimally from that point on. The Q-function can be learned iteratively by _____.

(A) The immediate reward plus discounted maximum future reward, Bellman equation

(B) Discounted maximum future reward, Markov decision process

(C) The immediate reward plus discounted maximum future reward, Markov decision process

(D) Discounted maximum future reward, Bellman equation

14) When we use one deep convolutional neural network model to classify 101 concepts, which option is **not** correct in the following description?_____

(A) The output of the last fully connected layer can be used as the learning features of each concept.

(B) The dimension of the classification layer can be 101.

(C) The convolutional kernels are pre-defined (i.e., data-independent).

(D) Dropout is used to boost the performance.

15) Which description is **not** correct in terms of deep learning?_____

(A) Deep learning is essentially a method to learn the features of raw data.

(B) Backpropagation is conducted to optimize the weights of deep neural networks so that the neural network can learn how to correctly map arbitrary inputs to outputs.

(C) The achieved performance of deep learning is due to its powerful representation ability via many of non-linear mappings.

(D) Deep convolutional neural network for classification is employed in an end-to-end mechanism via unsupervised learning.

16) Which description is **not** correct about K-Means clustering?_____

(A) K-means clustering can be used for image segmentation and image compression.

(B) K is the number of clusters and is generally pre-defined.

(C) Each data point can be assigned to more than one cluster.

(D) If the dimension of each data points is D, the dimension of cluster centers is D.

17) The number of pruned successors in alpha-beta pruning is highly dependent on _____.

(A) The moving order

(B) The initialized values of alpha and beta

(C) The number of terminal nodes

(D) Whether breadth-first search or depth-first search is employed

18) What description is **not** correct in terms of AI?

(A) Deep learning is one kind of machine learning methods.

(B) Machine learning is deep learning.

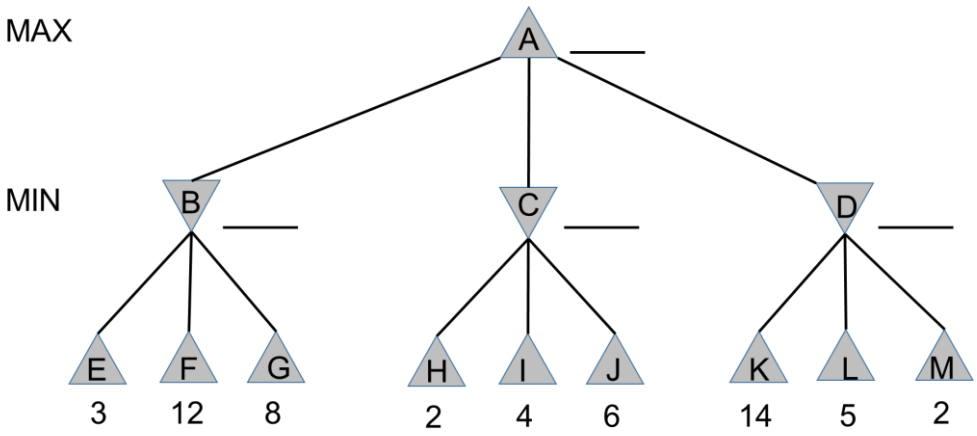
(C) Search is one kind of methods used in AI.

(D) In general, LeNet-5 (one of deep convolutional neural networks) maps each handwriting images into 0-9 digital character concept space.

3. Calculus and Analysis (34 points)

1) (**Game Playing, 8 points**) As shown in the following figure, there is a MINMAX search tree with three layers. The utility values for the leaf nodes are respectively displayed at the bottom of the figure.

(a) Fill in the blanks for the utility values associated with the tree nodes (i.e., B, C, D) as well as the root node A. (4 points)



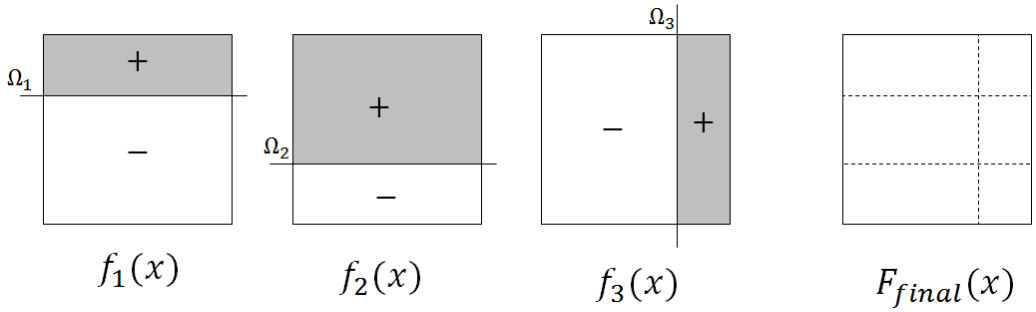
(b) Draw mark '//' on some branches in the figure to show that they are pruned by the alpha-beta pruning algorithm. (4 points)

2) (**Boosting, 8 points**) Boosting is a powerful technique for combining multiple "base" classifiers to produce a form of committee whose performance can be significantly better than that of any of the base classifiers. Consider a two-class classification problem, in which the training data comprises 2D input vectors $\mathbf{x}_1,...,\mathbf{x}_N$ along with corresponding binary target variables $t_1,...,t_N$ where $t_n \in \{+1,-1\}$. Assume that we have trained three base classifiers $f_1(\mathbf{x}), f_2(\mathbf{x}), f_3(\mathbf{x})$ and the corresponding weighting coefficients $\alpha_1,\alpha_2,\alpha_3$. Please answer:

(a) The final classifier learned by boosting can be given by: (4 points)

$F_{final}(\mathbf{x}) =$

(b) If three base classifiers $f_1(\mathbf{x}), f_2(\mathbf{x}), f_3(\mathbf{x})$ are shown in the figure and $\alpha_1 = 0.3, \alpha_2 = 0.5, \alpha_3 = 0.7$. Each base classifier partitions the input space into two regions separated by a linear decision boundary Ω_i. The dark region with '+' means the target value is +1 and the bright region with '-' means the target value is -1. Three decision boundaries have been put together in the last right figure and separate the space into six sub-regions, please mark the final decision result in the figure with '+' or '-' for each sub-region.(4 points)



3) (**Image Restoration, 6 points**) Please share several key tricks that effectively improve the performance of your Image Restoration Algorithm in Project 2. (About 100~150 words).

4) (**Deep learning, 12 points**)

(a) Convolution is very important in deep convolutional neural network. Please calculate the convolved value of the center pixel in Figure (1) with the given convolutional kernel in Figure (2). (3 points)

15	10	10	10	10
10	9	7	9	9
10	9	**8 (center pixel)**	9	8
12	8	7	8	7
11	8	7	8	6

Figure (1)

1	-1	0
1	-1	1
0	-1	1

Figure (2)

(b) Given a single depth slice in Figure (3), please give out the average-pooling value of this slice with 2×2 filters and stride 2. (3 points)

8	8	6	6
10	6	8	4
10	10	9	7
10	10	9	7

Figure (3)

(c) If we trained a deep convolutional neural network as follows in Figue (4). The sofmax is used to classify five concepts (e.g., car, airplane, truck, ship and person). If we input a car image into the trained deep model, please write out one of likely 5-dimensional outputs by the deep model. (3 points)

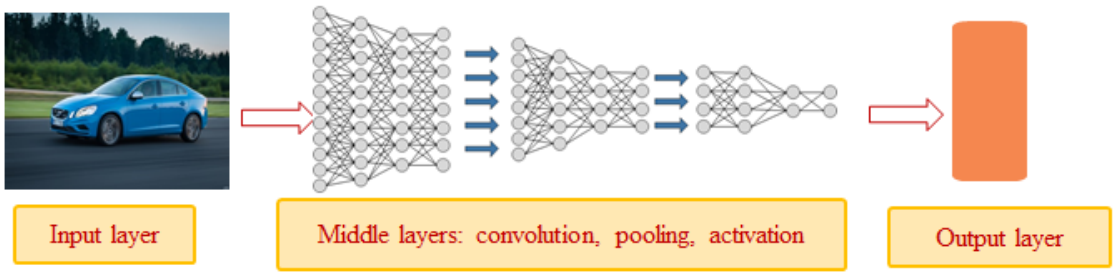


Figure (4)

(d) Please write down the trainable parameters in this model. (3 points)