CS540 ANSWER SHEET

Name _____ Email _____

1.	2.
3.	4.
5.	6.
7.	8.
9.	10.
11.	12.
13.	14.
15.	16.
17.	18.
19.	20.

Midterm Examination CS540-1: Introduction to Artificial Intelligence Fall 2014
20 questions, 5 points each

INSTRUCTIONS: Choose ONE answer per question. WRITE YOUR ANSWERS ON THE ANSWER SHEET. WE WILL NOT GRADE ANSWERS ON OTHER PAGES. ON THE ANSWER SHEET WRITE DOWN THE ANSWERS ONLY – DO NOT INCLUDE INTERMEDIATE STEPS OR DERIVATIONS. BE SURE TO INCLUDE YOUR NAME AND EMAIL ON THE ANSWER SHEET, TOO.

1. What would be the output tree of HAC using complete linkage on the following 1D points: -3, -1, 0, 4, 8, 10?

   (A) ((-3 (-1 0)) (4 8) 10)) (B) (-3 ((-1 0) (4 (8 10))))

   (*C) ((-3 (-1 0)) (4 (8 10))) (D) (((-3 (-1 0)) 4) (8 10)) (E) none of the above

2. Consider a binary classification task in $\mathbb{R}^2$ with the training data:

   Class P: (1,1), (2,1), (3,2), (-1, 3)

   Class N: (2,3), (3,3), (-1,-1)

   Use 3-NN with Euclidean distance. What's the label for test point (2,2), before and after adding another class N training point (2.2, 1.8) to the training set?

   (A) P, P (B) P, N (C) N, P (D) N, N (*E) Ambiguous because of ties.

3. Let $f()$ be a continuous differentiable function in $\mathbb{R}$. If the derivative $f'(x) = 0$ at x, which one of the following is impossible at x?

   (A) local minimum (B) local maximum (C) global minimum

   (D) global maximum (*E) Gradient descent on $f()$ starting at x will move away.

4. Which function is convex on $\mathbb{R}$?

   (A) $-x^2$ (*B) $-x + 2$ (C) x^3 (D) $\log_2(x^2)$ (E) none of the above

5. Let k-means clustering starts with the centroids (2,3), (4,1). The data points are (7,10), (0,-2), (-3,4), (2,5). After one iteration of k-means (assigning the points to clusters, and updating the centroids), what is the distortion?

   (A) 2 (B) 24 (C) 46 (D) 68 (*E) none of the above

6. In $\mathbb{R}^4$ what is the city-block distance (also known as $L1$ distance or Manhattan distance) between two points $(1, 2, 3, 4)$ and $(8, 7, 6, 5)$?

   (A) 9 (B) 9.1652 (*C) 16 (D) 16.1652 (E) none of the above

7. A test set $(x_1, y_1), \ldots (x_{1000}, y_{1000})$ contains labels $y_i = (-1)^i$ for $i = 1 \ldots 1000$. A classifier simply predicts $y = 1$ all the time. What is this classifier's test accuracy?

   (A) 0 (B) 0.001 (*C) 0.5 (D) 1 (E) none of the above

8. Which classifier can never produce a linear decision boundary in $\mathbb{R}^2$?

   (A) kNN (B) decision tree (C) SVM (D) neural network

   (*E) They all can.

9. A decision tree has depth d (a decision tree where the root is a leaf node has $d = 1$). All its internal node has c children. The tree is also complete, meaning all leaf nodes are at depth d. If we require each leaf node to contain at least 15 training examples, what is the minimum size of the training set?

   (*A) $15c^{d-1}$ (B) $15c^d$ (C) $15(1 + c + c^2 + \ldots + c^{d-1})$

   (D) $15(1 + c + c^2 + \ldots + c^d)$ (E) none of the above

10. A bag contains a red ball, a green ball, and a blue ball. Randomly draw a ball from the bag with equal probability. What is the entropy of the outcome?

    (A) $\log_2(2)$ (*B) $\log_2(3)$ (C) $\log_3(2)$ (D) $\log_3(3)$ (E) none of the above

11. If a feature (i.e. attribute) and the class label are independent (in a probability sense), what is the mutual information between them?

    (A) negative (*B) 0 (C) positive (D) can be anything (E) none of the above

12. A linear SVM in $\mathbb{R}^5$ has $w = (1, 0, 1, 0, 1)^\top$ and $b = -10$. What class does it predict on test point $x = (5, 4, 0, 1, 0)$?

    (*A) -1 (B) 1 (C) -5 (D) 5 (E) none of the above

13. A linear SVM with slack variables has the objective function $\frac{1}{2}w^\top w + C\sum_i \xi_i$. Which statement is FALSE when $C = 0$?

    (*A) The resulting decision boundary can correctly separate all training points.

    (B) The slack variables ξ_i cannot be negative.

    (C) There is always a solution w.

    (D) All inequality constraints can be satisfied.

    (E) It is possible for a training point to be on the decision boundary.

14. Suppose an SVM has $w = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$ and $b = -1$. What is the actual distance between the two planes defined by $w^\top x + b = -1$ and $w^\top x + b = 1$?

(A) $\frac{1}{\sqrt{14}}$ (*B) $\frac{2}{\sqrt{14}}$ (C) $\frac{\sqrt{11}}{2}$ (D) $\sqrt{11}$ (E) none of the above

15. If $K(x, x')$ is a kernel with induced feature representation $\phi(x)$, and $G(x, x')$ is another kernel with induced feature representation $\theta(x)$, then it is known that $H(x, x') = aK(x, x') + bG(x, x')$ is also a kernel if $a > 0, b > 0$. What is the induced feature representation of H?

(A) $\sqrt{a}\phi(x)+\sqrt{b}\theta(x)$ (B) $a\phi(x)+b\theta(x)$ (C) $\begin{pmatrix} a\phi(x) \\ b\theta(x) \end{pmatrix}$ (*D) $\begin{pmatrix} \sqrt{a}\phi(x) \\ \sqrt{b}\theta(x) \end{pmatrix}$

(E) none of the above

16. Due to some bizarre reason one of the inputs to a sigmoid perceptron is positive infinity! What output value is possible for this perceptron?

(A) 0 (B) 0.5 (C) 0.540 (D) 1 (*E) all of the above

17. For gradient descent to find the minimum of the function $f(x) = 0.25x^2$, which step size α ensures convergence?

(A) 0 (*B) 2 (C) 4 (D) 6 (E) none of the above

18. Which type of classifier cannot handle multiclass classification problems?

(A) kNN (B) decision tree (C) SVM (D) neural network (*E) They all can.

19. Bag A contains 5 white balls and 7 black balls. Bag B contains 3 whites and 12 black. A fair coin is flipped; if it is Heads, a ball is drawn from bag A, and if it is Tails, a ball is drawn from bag B. Suppose that this experiment was done and a white ball was drawn. What is the probability that this ball was in fact taken from bag B? (i.e. that the coin flip was Tails.)

(A) $\frac{17}{37}$ (*B) $\frac{12}{37}$ (C) $\frac{17}{35}$ (D) $\frac{12}{35}$ (E) none of the above

20. Which of the following is impossible?

(A) $P(A \mid B) > P(A)$ (B) $P(A \mid B) < P(A)$ (*C) $P(A, B) > P(A)$
(D) $P(A, B) < P(A)$ (E) $P(A, B) = P(A \mid B)$