CS540 Final Exam

December 14

Fall 2024

1 Student Information

- Exam version: 1
- First, Last Name:
- 10-digit Campus ID:
- Section: $\boxed{001}$ MW $4:00$, $\boxed{002}$ TR $2:30$, $\boxed{003}$ TR $1:00$
- Net ID (Wisc Email ID):

2 Instructions

1. You cannot sit next (front, back, left (within two seats including an empty seat), right (within two seats including an empty seat)) to someone you know. Switch seats now if that is the case.
2. Fill in these fields (left to right) on the Scantron sheet using pencil:
   - LAST NAME (family name) and FIRST NAME (given name), fill in bubbles
   - IDENTIFICATION NUMBER is your Campus ID number, fill in bubbles
   - Under A, B, C of SPECIAL CODES, fill in your three-digit section number $(001,$

   $002, 003)$

   - Under D of SPECIAL CODES, fill in Exam version above. This is very important!

   Make sure you fill all the above accurately in order to get graded.
3. Mark your answers on the Scantron sheet and this handout. Use a pencil to mark all answers. Each question has exactly one correct answer.
4. You may only reference your one-sheet notes. Please turn off and put away portable electronics now.
5. If you need to use the restroom, your phone must remain in the room, placed face down on your desk and kept visible to the proctors.
6. You have 120 minutes to take the exam. Once you have finished, please show your student ID to one of the proctors and submit both your Scantron and this handout. You do not need to submit your one-sheet note.

Good luck!

-

3 Questions

1. Consider the problem of detecting if an email message is a spam. Three variables are used to model this problem: a binary label S indicates if the message is a spam, and two binary features: C, F indicating whether the message contains "Cash" and "Free". A Naive Bayes classifier is trained with the following estimated probabilities:

$$\mathbb{P}\{S = 1\} = 0.5$$
$$\mathbb{P}\{C = 1|S = 0\} = 0.3, \mathbb{P}\{F = 1|S = 0\} = 0.2$$
$$\mathbb{P}\{C = 1|S = 1\} = 0.8, \mathbb{P}\{F = 1|S = 1\} = 0.9$$

What is the probability that the email is a spam given it does not contain "Cash" or "Free", that is $\mathbb{P}\{S = 1|C = F = 0\}$? You can assume $\mathbb{P}\{C = F = 0\} = 0.29$.

- A: $\dfrac{0.5 \cdot 0.7 \cdot 0.8}{0.29}$
- B: $1 - \dfrac{0.5 \cdot 0.8 \cdot 0.9}{0.29}$
- C: $1 - \dfrac{0.5 \cdot 0.3 \cdot 0.2}{0.29}$
- D: $\dfrac{0.5 \cdot 0.2 \cdot 0.1}{0.29}$
- E: None of the above (or more information is needed)

2. Suppose eigenfaces are computed based on a dataset containing 100 images, each 8 pixels by 8 pixels. The first 6 eigenvectors of the covariance matrix are used as eigenfaces. Which of the following about one eigenface is true?

- A: Each eigenface is a vector with 48 elements
- B: Each eigenface is 64 pixels in total
- C: Each eigenface is 36 pixels in total
- D: Each eigenface is a vector with 100 elements
- E: None of the above (or more information is needed)

3. The pairwise distance matrix between the four clusters $\{1, 2, 3, 4\}$ is given by $\begin{bmatrix} 0 & 1 & 2 & 3 \\ 1 & 0 & 4 & 5 \\ 2 & 4 & 0 & 6 \\ 3 & 5 & 6 & 0 \end{bmatrix}$. Which two clusters are merged in the next iteration of hierarchical clustering?

- A: $\{3, 4\}$ using single linkage, $\{1, 2\}$ using complete linkage
- B: $\{1, 2\}$ using single linkage, $\{3, 4\}$ using complete linkage
- C: $\{1, 2\}$ using both single and complete linkage distances
- D: $\{3, 4\}$ using both single and complete linkage distances
- E: None of the above (or more information is needed)

4. You are given a training set of five points (with 1 feature) and their 2-class classifications ($+$ or $-$) : $\{(-7, +), (-3, +), (-1, -), (1, -), (3, -)\}$. With this training set, the 3NN (3 Nearest Neighbor) algorithm will classify a given feature x as $+$ if $x < t$ and $-$ if $x > t$. What is t?

- A: -4
- B: 1
- C: -2
- D: -3
- E: None of the above (or more information is needed)

5. Consider the hard margin support vector machine classifier applied to a binary classification problem with two features. Suppose there are only three support vectors $\begin{bmatrix} 0 \\ 2 \end{bmatrix}$ with label 0, $\begin{bmatrix} 2 \\ 0 \end{bmatrix}$ with label 0, and $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ with label 1. What is the margin (thickness of the thickest line that can separate the two classes) of this hard margin support vector machine classifier?

   - A: $\sqrt{2}$
   - B: $2\sqrt{2}$
   - C: 1
   - D: 2
   - E: None of the above (or more information is needed)

6. Suppose you are given a fully connected neural network with 2 hidden layers, 2 input units, 4 hidden units in the first hidden layer, 5 hidden units in the second hidden layers, and 3 output units. How many weights and biases are there in this neural network?

   - A: $43 + 14$
   - B: $43 + 12$
   - C: $38 + 12$
   - D: $38 + 14$
   - E: None of the above (or more information is needed)

7. In a three-layer fully connected neural network, for training item i, let $a_{i2}^{(1)}$ be the second tanh activation unit in the first hidden layer and $a_{i3}^{(2)}$ be the third sigmoid activation unit in the second hidden layer. What is $\dfrac{\partial a_{i3}^{(2)}}{\partial a_{i2}^{(1)}}$? For tanh activation function $g(z)$, the derivative $g'(z) = 1 - g(z)^2$, and for sigmoid activation function $g(z)$, the derivative $g'(z) = g(z)(1 - g(z))$. Note: $w_{23}^{(2)}$ is the weight connecting unit 2 in the first hidden layer and unit 3 in the second hidden layer and $w_{32}^{(2)}$ is the weight connecting unit 3 in the first hidden layer and unit 2 in the second hidden layer.

   - A: $\left(1 - \left(a_{i3}^{(2)}\right)^2\right) w_{23}^{(2)}$
   - B: $a_{i3}^{(2)}\left(1 - a_{i3}^{(2)}\right) w_{23}^{(2)}$
   - C: $a_{i3}^{(2)}\left(1 - a_{i3}^{(2)}\right) w_{32}^{(2)}$
   - D: $\left(1 - \left(a_{i3}^{(2)}\right)^2\right) w_{32}^{(2)}$

- E: None of the above (or more information is needed)

8. In one step of gradient descent for a L_2 regularized logistic regression, suppose $w = 1, b = -1$, and $\dfrac{\partial C}{\partial w} = 2, \dfrac{\partial C}{\partial b} = -2$. The learning rate is $\alpha = 1$ and the regularization parameter $\lambda = \dfrac{1}{2}$, what is the weight w after one iteration? Note: the total loss is $C\left(w, b\right) + \dfrac{\lambda}{2}\left\|\begin{bmatrix} w \\ b \end{bmatrix}\right\|_2^2 = C\left(w, b\right) + \dfrac{\lambda}{2}\left(w^2 + b^2\right)$.

   - A: 3.5
   - B: -1.5
   - C: -1.25
   - D: 3.25
   - E: None of the above (or more information is needed)

9. What is the convolution between the image $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ and the filter $\begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ (same when flipped)? Assume zero padding and a stride of 1.

   - A: $\begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}$
   - B: $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
   - C: $\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$
   - D: $\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$
   - E: None of the above (or more information is needed)

10. Consider a recurrent neural network with 4 input features, 2 hidden recurrent units, and 1 output unit at the end of the sequence. During one step of stochastic gradient descent, for one training item, a sequence of length 7, the values of 14 weights are updated. How many weights will be updated for another training sequence of length 8? Note: a weight is considered updated if one gradient descent step is applied, including when the gradient has value 0.

    - A: 14
    - B: 15
    - C: 18
    - D: 16
    - E: None of the above (or more information is needed)

11. Consider a search problem with states $\{0, 1, 2, 3, 4, ...\}$. The initial state is 1, and the goal state is 4. State 0 has no successor and states $i \in \{1, 2, 3, 4, ...\}$ each has two successors $\{i + 1, 0\}$. We do not keep track of which states are checked previously, so we may expand the same state multiple times. How many states (including repeated ones) will be expanded by BFS (Breadth First Search)? Break ties by expanding the state with the larger index first.

- A: 8
- B: 6
- C: 7
- D: 9
- E: None of the above (or more information is needed)

12. Consider the same search problem as in the previous question. How many states (including repeated ones) will be expanded by DFS (Depth First Search)? Break ties by expanding the state with the larger index first.

- A: 9
- B: 5
- C: 8
- D: 4
- E: None of the above (or more information is needed)

13. Consider the same search problem as in the previous question. How many states (including repeated ones) will be expanded by IDS (Iterative Deepening Search)? Break ties by expanding the state with the larger index first.

- A: $1 + 3 + 5 + 7$
- B: $1 + 3 + 5 + 6$
- C: $1 + 3 + 5 + 4$
- D: $1 + 3 + 5 + 5$
- E: None of the above (or more information is needed)

14. Suppose the states are integers between 1 and 1023. The initial state is 1, and the goal state is 1023. The successors of state i are $2i$ and $2i + 1$, if exist. What is the largest number of states on the frontier (or fringe) at the same time if BFS (Breadth First Search) is used? Break ties by expanding the state with the smaller index first. Note: $1024 = 2^{10}$ and $512 = 2^{9}$.

- A: 512
- B: 11
- C: 1024
- D: 12
- E: None of the above (or more information is needed)

15. Consider the same search problem as in the previous question with the same tie breaking rule. What is the largest number of states on the frontier (or fringe) at the same time if IDS (Iterative Deepening Search) is used?

- A: 512
- B: 12

- C: 11
- D: 1024
- E: None of the above (or more information is needed)

16. Suppose there are four states in the priority queue during a search, $(s = 1, g = 3, h = 2)$, $(s = 2, g = 1, h = 3)$, $(s = 3, g = 4, h = 1)$, and another state $s = 4$. Here, g is the cost so far (from the initial state to the current state) and h is the heuristic cost. Suppose that A search ($A^\star$ without the $\star$) is used, and that state $s = 4$ is expanded next. What are the possible values of g and h for the state $s = 4$? Break ties by expanding the state with the largest state index first.

    - A: $(s = 4, g = 2, h = 2)$ is possible, $(s = 4, g = 3, h = 1)$ is not possible.
    - B: $(s = 4, g = 2, h = 2)$ and $(s = 4, g = 3, h = 1)$ are both possible.
    - C: $(s = 4, g = 3, h = 1)$ is possible, $(s = 4, g = 2, h = 2)$ is not possible.
    - D: $(s = 4, g = 2, h = 2)$ and $(s = 4, g = 3, h = 1)$ are both not possible.
    - E: None of the above (or more information is needed)

17. Consider a search graph with 10 states $\{1, 2, 3, ..., 10\}$. For every $1 \leqslant i < j \leqslant 10$, there is a directed edge from i to j with an edge cost i. The initial state is 1 and the goal state is 10. How many unique states are expanded if UCS (Uniform Cost Search) is performed (include the initial state and the goal state)? Break ties by expanding the state with the larger index.

    - A: 10
    - B: 3
    - C: 9
    - D: 2
    - E: None of the above (or more information is needed)

18. Consider the same search problem as in the previous question, which one of the following heuristic functions is admissible? For $s \in \{1, 2, 3, ..., 10\}$,

    - A: $h(s) = \frac{s}{10}$
    - B: $h(s) = 1 - \frac{s}{5}$
    - C: $h(s) = 1$
    - D: $h(s) = 1 - \frac{s}{10}$
    - E: None of the above (or more information is needed)

19. Consider a different search graph with 10 states $\{1, 2, 3, ..., 10\}$. For every $1 \leqslant i < j \leqslant 10$, there is a directed edge from i to j with an edge cost j. The initial state is 1 and the goal state is 10. The admissible heuristic $h(s) = 10 - s$ is used. How many unique states are expanded if GS (best first Greedy Search) is performed (including the initial state and the goal state)? Break ties by expanding the state with the larger index.

    - A: 10

- B: 9
- C: 3
- D: 2
- E: None of the above (or more information is needed)

20. Consider the same search problem as in the previous question, which one of the following heuristic is also admissible and dominates the $h(s) = 10 - s$ heuristic from the previous question? For $s \in \{1, 2, 3, ..., 10\}$,

    - A: $h(s) = \max\{9 - s, 0\}$
    - B: $h(s) = 10$
    - C: $h(s) = 11 - s$
    - D: $h(s) = 10 \min\{10 - s, 1\}$
    - E: None of the above (or more information is needed)

21. There are 45 cookies. The brother first proposes a division of these cookies into two piles (two integers adding up to 45), then the sister take one of the two piles. Both the brother and the sister want to maximize the number of cookies they take. What is an optimal action of the brother (action in the solution of the game)?

    - A: $(1, 44)$
    - B: $(22, 23)$
    - C: $(15, 30)$
    - D: $(0, 45)$
    - E: None of the above (or more information is needed)

22. Two countries simultaneously choose one of two actions: country A chooses between attack country B or not attack, and country B chooses to fight or flee. The rewards (country A, country B) from each pair of actions are summarized in the table: $\begin{bmatrix} A \backslash B & \text{fight} & \text{flee} \\ \text{attack} & (-1, -1) & (1, 0) \\ \text{not} & (0, 0) & (0, 0) \end{bmatrix}$ What is the complete set of pure strategy Nash equilibria of this game?

    - A: Only $\{(\text{ attack }, \text{ flee })\}$
    - B: $\{(\text{ attack }, \text{ flee }), (\text{ not }, \text{ fight })\}$
    - C: Only $\{(\text{ not }, \text{ fight })\}$
    - D: There are no pure strategy Nash equilibria
    - E: None of the above (or more information is needed)

23. Consider the sequential version of the game in the previous question with the same reward table, where country A chooses between attack or not attack, and then if country A chooses to attack, country B can choose between fight and flee; but if country A chooses not to attack, the game ends (and both countries get 0). What is the complete set of solutions (that is, the optimal actions for countries A and B) of this game?

- A: Only $\{($ not , fight $)\}$
- B: Only $\{($ attack , flee $)\}$
- C: There are no solutions
- D: $\{($ attack , flee $),($ not , fight $)\}$
- E: None of the above (or more information is needed)

24. Consider a zero-sum sequential game where the max player first chooses one of two actions U and D, and then the min player chooses one of three actions L, C, and R. The rewards are summarized in the table $\begin{bmatrix} \max \backslash \min & L & C & R \\ U & 6 & 4 & 3 \\ D & 5 & 2 & 1 \end{bmatrix}$. The order of how the states will be expanded is $U \rightarrow D$ and $L \rightarrow C \rightarrow R$. How many actions will be pruned during $\alpha - \beta$ pruning? In the case when $\alpha = \beta$, the remaining actions will be pruned.

- A: 4
- B: 1
- C: 2
- D: 0
- E: None of the above (or more information is needed)

25. Consider the same game as in the previous question, and suppose the static evaluation function (or heuristic function) is $h(U) = 0$ and $h(D) = 1$. Which action will the max player choose with and without using the heuristic? Note: the static evaluation function outputs value for the max player.

- A: U if the heuristic is used, and D if the heuristic is not used
- B: D if the heuristic is used, and U if the heuristic is not used
- C: U with or without using the heuristic
- D: D with or without using the heuristic
- E: None of the above (or more information is needed)

26. Consider a zero-sum sequential game where the max player chooses one of 5 actions, and after the max player chooses any action, the min player chooses one of 4 actions. During the Alpha Beta pruning process, what is the maximum number of actions of the min player that can be pruned (assuming the actions can be reordered)?

- A: 16
- B: 9
- C: 15
- D: 12
- E: None of the above (or more information is needed)

27. What is the complete list of rationalizable actions (actions that survive the iterative elimination of strictly dominated actions) for the zero-sum normal-form game $\begin{bmatrix} \max \setminus \min & L & C & R \\ U & 1 & 0 & -1 \\ M & 0 & 1 & -1 \\ D & -1 & -1 & -1 \end{bmatrix}$? Note: the values in the matrix is for the max player.

- A: $\{U, M\}$ and $\{R\}$
- B: $\{U, M\}$ and $\{L, C\}$
- C: $\{U, M, D\}$ and $\{L, C, R\}$
- D: $\{D\}$ and $\{R\}$
- E: None of the above (or more information is needed)

28. Two players simultaneously offer x_1 and x_2 dollars to buy 10 dollars from the auctioneer. The player who offers the higher dollar amount will receive the 10 dollars and the other player will receive 0 dollars. In case of tie, each player will get 5 dollars. Each players' rewards are given by the money received minus the money offered, in particular, the player with the lower offer will get a reward equal to the negative of what they offered. Which one of the following values of (x_1, x_2) is a Nash equilibrium of this game?

- A: $(10, 0)$
- B: $(10, 10)$
- C: $(0, 0)$
- D: $(5, 5)$
- E: None of the above (or more information is needed)

29. The BoS (Battle of Sexes) game $\begin{bmatrix} \text{row} \setminus \text{col} & B & S \\ B & (x, y) & (0, 0) \\ S & (0, 0) & (1, 1) \end{bmatrix}$ has a Nash equilibrium where the row player uses B with probably $\frac{1}{3}$ and the column player uses B with probability $\frac{1}{2}$. What are the values of (x, y)?

- A: $(1, 2)$
- B: $(2, 1)$
- C: $\left(1, \frac{1}{2}\right)$
- D: $\left(\frac{1}{2}, 1\right)$
- E: None of the above (or more information is needed)

30. Consider a simultaneous move game where the column player can choose between "cheat" (C) and "not cheat" (N), and the row player can choose between "investigate" (I) and "drop investigation" (D). The rewards are given by $\begin{bmatrix} \text{row} \setminus \text{col} & C & N \\ I & (-2 + r, -4) & (-2, 0) \\ D & (-r, 1) & (-r, 0) \end{bmatrix}$. What is the smallest value of r such that (I, N) is a pure strategy Nash equilibrium of this game?

- A: 2
- B: -2
- C: 0
- D: 1
- E: None of the above (or more information is needed)

31. Suppose the UCB1 (Upper Confidence Bound) algorithm is used to select arms in a multi-armed bandit problem, and in round 12, arm 1 is pulled 3 times and the empirical mean is 3, arm 2 is pulled 5 times and the empirical mean is 2, arm 3 is pulled 4 times and the empirical mean is 1. Recall that the upper confidence bound at round t for arm k is calculated as $u_k = \hat{\mu}_k + c\sqrt{\frac{2\log(t)}{n_k}}$, where μ_k is the empirical mean and n_k is the number of rounds in which arm k is pulled. Which one of the following statements about u_1, u_2, u_3 is correct?

    - A: $u_1 > u_2$ and $u_1 > u_3$ and comparison between u_2 and u_3 depends on c
    - B: $u_1 > u_3$ and comparison with u_2 depends on c
    - C: $u_1 > u_2 > u_3$
    - D: $u_1 > u_2$ and comparison with u_3 depends on c
    - E: None of the above (or more information is needed)

32. Consider the MDP (Markov Decision Process) with two states 1 and 2, and actions "stay" (s) and "move" (m). Staying in state 1 gives reward 0.8 and staying in state 2 gives reward 1. Moving gives reward 0. With discount factor $\beta = 0.75$, what is the value of the optimal policy in state 1 (that is, what is $V^\star(1)$)?

    - A: 3
    - B: 4
    - C: 1
    - D: 3.2
    - E: None of the above (or more information is needed)

33. Consider the same MDP as in the previous question. What is the value of $Q^\star(1, \text{ move })$ under the optimal policy?

    - A: 1.8
    - B: 3.2
    - C: 3
    - D: 2.4
    - E: None of the above (or more information is needed)

34. Consider the same MDP as in the previous question, but let the reward of moving (in both directions) be r where $0 \leqslant r \leqslant 1$. What is the maximum possible value of r such that "stay" an optimal action in state 1? Note: there could be more than one optimal policies.

- A: 0.2
- B: 0
- C: 1
- D: 0.6
- E: None of the above (or more information is needed)

35. There are 3 states $\{1, 2, 3\}$ and 3 actions $\{1, 2, 3\}$. During one iteration of Q-learning, we start from state 2, choose action 2 and receive reward 1, and in the next iteration, we will end up in state 3 and choose action 3. The current Q table is given by $\begin{bmatrix} s \backslash a & a=1 & a=2 & a=3 \\ s=1 & 2 & 4 & 6 \\ s=2 & -2 & -4 & -6 \\ s=3 & 2 & 0 & -2 \end{bmatrix}$. What is the updated value of $Q\left(s=2, a=2\right)$? Use the learning rate $\alpha = 1$ and the discount rate $\beta = 0.5$.

- A: -3
- B: 2
- C: 0
- D: -5
- E: None of the above (or more information is needed)

36. Consider the same setting as in the previous question but SARSA is used instead of Q-learning. What is the updated value of $Q\left(s=2, a=2\right)$?

- A: -3
- B: 0
- C: -5
- D: 2
- E: None of the above (or more information is needed)

37. Suppose the hill climbing algorithm is applied to a problem with 10 states $\{1, 2, ..., 10\}$, where state i has two neighbors $i-1$ and $i+1$, if exist. The score of state i is $\left(-1\right)^{i} i^{2}$, and we want to maximize the score. From which initial states will hill climbing reach the global maximum after finite number of iterations (that is, it is not stuck in a local maximum)?

- A: Only $\{10\}$
- B: Only $\{8, 9, 10\}$
- C: All of $\{1, 2, ..., 10\}$
- D: Only $\{9, 10\}$
- E: None of the above (or more information is needed)

38. We use simulated annealing in a score maximization problem, where $f\left(s\right)$ denotes the score of state s. Recall that if state t is superior to state s, we will move from s to t ; and if state t is inferior to state s , we will move from s to t with probability $e^{-\dfrac{1}{T}\left|f(s)-f(t)\right|}$. Suppose that $f\left(s\right) = 4$ and $f\left(t\right) = 9$, with $T = 5$. If we start from s, what is the probability we move to t?

- A: 1
- B: e^{-1}
- C: $1 - e^{-1}$
- D: e^{1}
- E: None of the above (or more information is needed)

39. When using the Genetic Algorithm, suppose the current population contains 4 states $\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$, $\begin{bmatrix} -1 \\ -1 \\ -1 \end{bmatrix}$, $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$, and $\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$. The fitness function is $f\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \left(x_1 x_2 x_3\right)^2$. What is the reproduction probability of the state with the highest fitness?

- A: $\dfrac{1}{2}$
- B: $\dfrac{2}{3}$
- C: $\dfrac{2}{5}$
- D: $\dfrac{3}{4}$
- E: None of the above (or more information is needed)

40. Consider the same population and fitness function from the previous question. Which one of the following state is impossible to be one of the states in the population in the next generation after single-point cross-over without mutation? The parents are chosen by sampling without replacement, meaning that reproduction between two copies of the same states is not allowed. If you think more than one of the four states are impossible, select E.

- A: $\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$
- B: $\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$
- C: $\begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}$
- D: $\begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$
- E: None or more than one of the above (or more information is needed)