ECE/CS 559 - Fall 2018 - Final.

FIRST, WRITE YOUR NAME AND ID BELOW.

Full Name: **ID Number:**

Each question is worth 4 points. In the matrix below, columns represent the question numbers, and rows represent the choices. Completely fill in the box corresponding to the question and its answer. Example: Suppose Q1 were a real question and your answer to Q1 were "D." Then, you would fill out the corresponding cell of the matrix as shown. Each question has only one correct answer. If you provide multiple answers to a question, you will not receive any credit for that question.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B
C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
■	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D
E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E

!!! YOU MUST MARK THE ANSWER SHEET ABOVE TO INDICATE YOUR ANSWER. YOUR SELECTIONS ON THE QUESTIONS THEMSELVES WILL NOT BE GRADED !!!

Q1: Attach a copy of the email confirming that you have completed the instructor & TA evaluations.

Q2: Consider solving the linear SVM for classes $\mathcal{C}_1 = \{\left[\begin{smallmatrix}1\\1\end{smallmatrix}\right], \left[\begin{smallmatrix}2\\1\end{smallmatrix}\right], \left[\begin{smallmatrix}1\\2\end{smallmatrix}\right]\}$ and $\mathcal{C}_2 = \{\left[\begin{smallmatrix}-1\\-1\end{smallmatrix}\right], \left[\begin{smallmatrix}-2\\0\end{smallmatrix}\right], \left[\begin{smallmatrix}0\\-2\end{smallmatrix}\right]\}$. What are the support vectors of the resulting solution?

a) $\{\left[\begin{smallmatrix}1\\1\end{smallmatrix}\right], \left[\begin{smallmatrix}-1\\-1\end{smallmatrix}\right]\}$ b) $\{\left[\begin{smallmatrix}1\\1\end{smallmatrix}\right]\} \cup \mathcal{C}_2$ c) $\{\left[\begin{smallmatrix}-1\\-1\end{smallmatrix}\right]\} \cup \mathcal{C}_1$ d) $\mathcal{C}_1 \cup \mathcal{C}_2$ e) $\{\left[\begin{smallmatrix}2\\1\end{smallmatrix}\right], \left[\begin{smallmatrix}1\\2\end{smallmatrix}\right], \left[\begin{smallmatrix}-2\\0\end{smallmatrix}\right], \left[\begin{smallmatrix}0\\-2\end{smallmatrix}\right]\}$

Q3: Which one of the following is the type of learning used when training a self-organizing map?
a) Unsupervised b) Supervised c) Reinforcement d) Federated e) Ensemble

Q4: Suppose that the goal is to minimize the function $f(x) = |x^2 + 3| - 1$ defined on $\mathbb{R}$. Then,
a) Newton's method will always converge to the global optimum in one iteration from any initial point.
b) Gradient descent will always converge to the global optimum in one iteration.
c) The problem is non-convex, so it is not feasible to find a solution.
d) All of a, b, and c are true. e) None of the claims in a, b, c, or d are true.

Q5: Suppose we compute the linear SVM on $n > 2$ data points in $\mathbb{R}^2$, yielding a hyperplane with exactly 2 support vectors. If we add one more data point and recompute the SVM, what is the maximum possible number of support vectors for the new hyperplane (assuming the $n + 1$ points are linearly separable)?
a) 1 b) 2 c) 3 d) n e) $n + 1$

Q6: Which one of the claims are true about the non-linear activation functions for neural networks?
I - They speed up the gradient calculation in backpropagation, as compared to linear units.
II - They help to learn nonlinear decision boundaries.
III - They are applied only to the output units.
a) Only I b) Only II c) I and II d) II and III e) I, II, and III

Q7: Which one of the following matrix properties does a kernel matrix always have?
I - Invertible II - Symmetric III - At least one negative eigenvalue IV - All the entries are positive
a) Only I b) Only II c) I and II d) I, II and III e) III and IV

Q8: Consider applying PCA to the set of vectors $\mathcal{C} = \{\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right], \left[\begin{smallmatrix}2\\0\end{smallmatrix}\right], \left[\begin{smallmatrix}3\\0\end{smallmatrix}\right], \left[\begin{smallmatrix}0\\1\end{smallmatrix}\right], \left[\begin{smallmatrix}0\\2\end{smallmatrix}\right]\}$. Which one of the below are the first and the second principal vectors, respectively?

a) $\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right], \left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]$ b) $\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right], \left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]$ c) $\frac{1}{\sqrt{2}}\left[\begin{smallmatrix}1\\1\end{smallmatrix}\right], \frac{1}{\sqrt{2}}\left[\begin{smallmatrix}-1\\1\end{smallmatrix}\right]$ d) $\frac{1}{\sqrt{2}}\left[\begin{smallmatrix}-1\\1\end{smallmatrix}\right], \frac{1}{\sqrt{2}}\left[\begin{smallmatrix}1\\-1\end{smallmatrix}\right]$ e) None of the other options.

Q9: Which one of the following best describes "the dropout method."

a) Random removal of edges and/or neurons in the network during training.

b) Intelligent modification of the learning rate to speed up convergence.

c) Removing biases/weights that are near zero to achieve a network with low computational complexity.

d) Removing training samples that result in large gradients from the training set. In this manner, one hopes to avoid divergence of weights during the training phase.

e) A 19th century progressive heavy metal band.

Q10: Which one of the following are correct statements about the backpropagation algorithm:

I - It is a learning algorithm for single or multilayer neural networks.

II - The backward pass follows after the forward pass.

III - It is based on the gradient descent method to minimize the classification errors.

a) Only I b) I and II c) I and III d) II and III e) I, II, and III

Q11: Which one of the following statements is correct regarding the perceptron training algorithm?

a) Using a small learning rate always leads to faster convergence.

b) Using a large learning rate always leads to faster convergence.

c) If the initial classes are not linearly separable, the weights do not converge.

d) It is a training algorithm for learning in multilayer neural networks.

e) It always yields the best separator that is also provided by the linear support vector machine.

Q12: How many layers of perceptrons are sufficient to implement any logic function of any number of inputs? Assume that the perceptrons use the step activation function $u(x)=1$, $x \geq 0$, and $u(x)=0$, $x < 0$.

a) 1 b) 2 c) 3 d) 4 e) One may need an arbitrarily large number of layers.

Q13: Consider an associative memory designed via the weight matrix $W = \sum_{i=1}^{n} x_i x_i^T$, where $x_1, \ldots, x_n \in \{-1, +1\}^d$ are the memory patterns that we wish to store. Consider a one layer feedforward network with weights W, zero biases, and the activation function $\operatorname{sgn}(x) = 1$, $x \geq 0$, and $\operatorname{sgn}(x) = 0$, $x < 0$. Note that a stored memory pattern z is one that satisfies $z = \operatorname{sgn}(Wz)$. Then, which one of the following is true?

a) Patterns z and $-z$ cannot simultaneously be stored memory patterns.

b) The maximum number of patterns that can be stored while guaranteeing perfect recovery for the stored patterns is exactly $0.18d$.

c) Restoration of a perturbed input to one of the stored memory patterns can be accomplished by using the idea of recurrence.

d) Any pattern that is not a stored memory pattern is called a spurious pattern.

e) With W designed as in the question, all $x_1, \ldots, x_n$ are always stored memory patterns.

Q14: Which one of the following statements are correct about unsupervised learning?

I- Algorithms for supervised learning are not directly applicable for unsupervised learning.

II- It is used when we have no information about the correct output of the training data.

III- The learning algorithm detects similarities between different training data inputs.

a) Only I b) Only II c) I and II d) II and III e) I, II, and III.

Q15: Consider applying the k-means algorithm to the set of vectors $\mathcal{C} = \left\{\left[\begin{smallmatrix}-1\\1\end{smallmatrix}\right], \left[\begin{smallmatrix}1\\1\end{smallmatrix}\right], \left[\begin{smallmatrix}1\\-1\end{smallmatrix}\right], \left[\begin{smallmatrix}-1\\-1\end{smallmatrix}\right]\right\}$ with initial centers $\left[\begin{smallmatrix}1\\0.5\end{smallmatrix}\right], \left[\begin{smallmatrix}-1\\-0.5\end{smallmatrix}\right]$. What are the resulting centers after the algorithm converges?

a) $\left[\begin{smallmatrix}1\\0.5\end{smallmatrix}\right], \left[\begin{smallmatrix}-1\\-0.5\end{smallmatrix}\right]$ b) $\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right], \left[\begin{smallmatrix}-1\\0\end{smallmatrix}\right]$ c) $\left[\begin{smallmatrix}1\\1\end{smallmatrix}\right], \left[\begin{smallmatrix}-1\\-1\end{smallmatrix}\right]$ d) $\left[\begin{smallmatrix}1\\-1\end{smallmatrix}\right], \left[\begin{smallmatrix}-1\\1\end{smallmatrix}\right]$ e) $\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right], \left[\begin{smallmatrix}0\\-1\end{smallmatrix}\right]$

Q16: Given $\alpha > 0$, what is the derivative of the sigmoidal activation $y = \frac{1}{1+e^{-\alpha x}}$ with respect to x?

a) $y(1-y)$ b) $\alpha y(1-y)$ c) y^2 d) αy^2 e) $\alpha y(1-y^2)$

Q17: Consider one layer of weights (edges) in a convolutional neural network (CNN) for images, connecting one layer of units to the next layer of units. Which type of layer has the fewest parameters to be learned during training?

a) A convolutional layer with ten 3×3 filters.
b) A convolutional layer with eight 5×5 filters.
c) A max-pooling layer that reduces a 10×10 image to a 5×5 image.
d) A fully-connected layer from 20 hidden units to 4 output units without biases.
e) A fully-connected layer from 4 hidden units to 20 output units with biases.

Q18: Suppose that you wish to regularize your cost function $f(w_1, w_2)$ using ℓ_1 regularization on the weights w_1, w_2. Which one of the following expressions would be the regularized cost function?
a) $f(w_1, w_2) + \lambda(|w_1|^2 + |w_2|^2)$ b) $f(w_1, w_2) + \lambda(|w_1| + |w_2|)$ c) $f(w_1, w_2) + \lambda\sqrt{|w_1|^2 + |w_2|^2}$.
a) $f(w_1, w_2) + \lambda(|w_1| + |w_2|)^2$ b) $f(w_1, w_2) + \lambda\sqrt{|w_1| + |w_2|}$

Q19: Consider a two-neuron Hopfield network with weight matrix $W = \left[\begin{smallmatrix} +2 & -1 \\ -1 & +0.5 \end{smallmatrix}\right]$, zero biases, and the $\mathrm{sgn}(\cdot)$ activation function; i.e., $\mathrm{sgn}(x) = 1$ if $x \geq 0$, and $\mathrm{sgn}(x) = -1$, otherwise. Consider the initial state $x = \left[\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right]$. What is the next state using the synchronous update rule.
a) $\left[\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right]$ b) $\left[\begin{smallmatrix} 1 \\ -1 \end{smallmatrix}\right]$ c) $\left[\begin{smallmatrix} -1 \\ 1 \end{smallmatrix}\right]$ d) $\left[\begin{smallmatrix} -1 \\ -1 \end{smallmatrix}\right]$ e) $\left[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right]$

Q20: Which one of the following are true about Hopfield networks?
I - The network states always converge with the asynchronous update rule given a symmetric weight matrix with non-negative diagonals.
II - It is impossible for all possible states of the network to be steady states.
III - It is impossible for all possible states of the network to be urstates.
a) Only I b) Only II c) Only III d) I and II e) I and III.

Q21: Which one of the following are true about competitive learning?
I - The neuron with the smallest induced local field or output is declared the winner neuron.
II - Typically, only the weights and the bias of the winning neuron are updated.
III - It is a special case of the Hebbian learning rule.
a) Only I b) Only II c) Only III d) II and III e) I, II and III.

Q22: Early stopping refers to stopping network training as soon as ...
a) the errors on the test set start increasing.
b) the errors on the training set start increasing.
c) the errors on the training set start decreasing.
d) the errors on the test set start decreasing.
e) the computer used for training starts overheating.

Q23: Oja's learning rule on a single perceptron is derived on the update $w \leftarrow w + \frac{w+\eta yx}{\|w+\eta xy\|}$ evaluated as $\eta \to 0$. Here, η is the learning parameter, $x \in \mathbb{R}^{d\times 1}$ is the perceptron input, and $y = w^T x$ is the perceptron output. According to this description, which one of the following is Oja's learning rule?
a) $w \leftarrow w + \eta y(x - w)$ b) $w \leftarrow w + \eta y(x - yw)$ c) $w \leftarrow w + \eta xy$
d) $w \leftarrow w + \eta x(\delta - y)$, where δ is the desired output. e) $w \leftarrow w + \eta x(\delta + y)$

Q24: Which one of the following are true about neural networks?
I- They can only be used to optimize convex cost functions.
II- They always output values between -1 and $+1$.
III- They can be used for classification as well as interpolation.
a) Only I b) Only II c) I and II d) I and III e) Only III

Q25: Consider a 3-layer convolutional neural network. The input is a 13×13 image. In the first layer, the 13×13 image is first mapped (without zero padding) to three 10×10 feature maps via three different 4×4 filters with stride 1. In the second layer, the three 10×10 feature maps are reduced to three 5×5 feature maps via max-pooling with stride 2. Finally, the three 5×5 feature maps are mapped to 4 outputs neurons via a fully connected layer. There are no biases in any one of the neurons. How many weights are there to optimize in this network?
a) 348 b) 160 c) 420 d) 128 e) 320