Name: Roll No:

Data Mining and Machine Learning Quiz 4, II Semester, 2024–2025 22 April, 2025

Some questions below may have more than one correct answer. You get full credit if you select all correct options. You get partial credit if you select a non-empty, strict subset of correct options. You get zero credit if you select any incorrect option.

- 1. Which of these statements are true of the perceptron algorithm?
 - (a) The algorithm always finds a separator if the data is linearly separable. \checkmark
 - (b) The time taken by the algorithm to converge depends on the magnitudes of the points in the dataset. \checkmark
 - (c) The perceptron algorithm is compatible with using kernels. \checkmark
 - (d) The perceptron algorithm terminates with a maximum margin separator.

Explanation:

(a) is the reason the perceptron algorithm converges. The proof uses the fact that the number of iterations is bounded by a function of the margin width and the magnitudes of the points in the dataset, so (b) is true. (c) follows from the fact that we can express the separator at each stage as a linear combination of the training data points.

(d) The perceptron algorithm terminates as soon as it finds a separator, which is not necessarily a maximum margin separator.

- 2. Which of these statements are true of support vector machines?
 - (a) When training soft-margin SVMs, if we allow a higher misclassification rate, we normally get a wider margin. \checkmark
 - (b) In the dual formulation solution, the values of the Lagrange multipliers identify the support vectors. \checkmark
 - (c) The dual formulation enables the use of kernel methods. \checkmark
 - (d) The dual formulation can be adapted to the soft margin case. \checkmark

Explanation:

(a) was demonstrated in the Python code for SVMs that was discussed in class. (b) is true because Lagrange multipliers are non-zero only for the support vectors. (c) follows from the fact that the dual formulation is entirely stated in terms of dot products. (d) follows from the fact that the soft margin case is also formulated as a quadratic optimization problem, which is solved using the dual.

- 3. Which of the following is/are true of the backpropagation algorithm?
 - (a) One of the inventors of the algorithm was awarded the Nobel Prize. \checkmark
 - (b) The gradient for weights in the later layers is likely to be smaller than those in initial layers.
 - (c) Backpropagation is not compatible with activation functions like RELU that are not everywhere differentiable.
 - (d) Backpropagation calculations can be speeded up through parallelization. $\checkmark.$

Explanation:

(a) Geoffrey Hinton was one of the winners of the Nobel Prize for Physics last year. (d) is true because updates to nodes within a layer can be done in parallel, though backpropagation proceeds sequentially through layers.

(b) is false because the gradients typically decrease when moving from the final layer to the earlier layers. (c) is empirically false because we saw that we can define networks with hidden layers that use RELU as the activation function.