

<b>Name:</b>	<b>Roll No:</b>	<b>Marks:</b>
--------------	-----------------	---------------

## Data Mining and Machine Learning

### Quiz 4, II Semester, 2025–2026

7 April, 2026

*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 about the perceptron algorithm is/are **not true**? (Mark the statements that you believe are **False**.)
  - (a) The algorithm always finds a separator if the data is linearly separable.
  - (b) The time taken by the algorithm to converge is directly proportional to the width of the margin. ✓
  - (c) The perceptron algorithm is not compatible with using kernels. ✓
  - (d) A network of perceptrons is capable of finding separators that a single perceptron cannot. ✓

*Explanation:*

(a) is true. We proved in class that the algorithm always converges if the data is linearly separable.

(b) is false. The narrower the margin, the larger the bound on the number of iterations before convergence, so the time taken is *inversely* proportional to the width of the margin.

(c) is false. The separator is always a linear combination of the training data points, so we only need to compute dot products between training data points and target points, allowing the use of kernel methods.

(d) is false. We argued that a network of perceptrons still computes only a linear function of the inputs.

2. Which of these statements is/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. ✓
  - (b) The dual formulation of the optimization problem enables the use of kernel methods. ✓
  - (c) We can algorithmically determine if a given function is a valid kernel.
  - (d) Expanding the input to include polynomial features (as we do in polynomial regression) is an alternative to using kernels. ✓

*Explanation:*

(a) was demonstrated in the Python code for SVMs that was discussed in class.

(b) follows from the fact that the dual formulation is entirely stated in terms of dot products.

(c) is not true. Mercer's Theorem is a mathematical characterization that is not effective. We have sufficient conditions and closure properties, but not a decision procedure.

(d) was demonstrated in the Python code for SVMs that was discussed in class.

3. Which of the following is/are true of the backpropagation algorithm? Assume that the network is layered and each layer is completely connected to the next layer.
  - (a) If we initialize two nodes in the same layer with the same set of weights and biases, they will have identical weights and biases after training. ✓
  - (b) There will always be multiple local minima equivalent to the one that we compute through gradient descent. ✓

- (c) Accuracy (the percentage of inputs classified correctly in a minibatch) can be used as a loss function.
- (d) Backpropagation calculations can be speeded up through parallelization. ✓.

*Explanation:*

(a) is true. Any two nodes in a layer have identical input and output connections. If the initial weights and biases are the same, they will evolve identically at every step of gradient descent and end up with the same final values.

(b) is true. If we permute the initial weights and biases within a layer, we will get a symmetric solution.

(c) is false. We need to be able to compute the loss function on individual inputs and compute the overall loss as an average of the individual losses.

(d) is true. Within a layer, we can update each node in parallel.

---