

Name: _____ Roll Number: _____

1. (a) (1 point) Gradient vector $\nabla_{\mathbf{w}}L(\mathbf{x}; \mathbf{w})$ lies in ...

- A. input domain \mathbf{x} .
- B. parameter space \mathbf{w} .
- C. output range of L .
- D. a direction orthogonal to L .

(b) (1 point) Gradient vector points in the direction of _____.

(c) (1 point) Describe the gradient descent method for minimizing a function.

(d) (1 point) What is the role of the learning rate η in gradient descent?

(e) Briefly describe the following.

i. (1 point) Batch Gradient Descent.

ii. (1 point) Stochastic Gradient Descent.

iii. (1 point) Stochastic Gradient Descent using Mini-Batches.

(f) (2 points) How does stochastic gradient descent help in avoiding local minima?

(g) (1 point) Near a minimum, if the step size is large, the gradient vector will start oscillating. How does the idea of momentum help in such a situation?