Name:	Roll No:	

Data Mining and Machine Learning

Quiz 2, II Semester, 2023–2024

4 April, 2024

1. To compute the parameters of an SVM, we move from the primal optimization problem

   - Minimize $\frac{|w|}{2}$
     Subject to $y_i \cdot (w_1 x_1^i + w_2 x_2^i + \cdots w_n x_n^i + b) > 1, i = 1, 2, \ldots, n$

   to the dual

   - Maximize $\sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} y_i y_j \alpha_i \alpha_j \langle x_i \cdot x_j \rangle$
     Subject to $\sum_{i=1}^{n} y_i \alpha_i = 0$ and $\alpha_i \geq 0, i = 1, 2, \ldots, n$

   What is the principal advantage of working with the dual formulation?

   (a) Computing the margin is more efficient.
   (b) Identifying the support vectors is easier.
   (c) The dual formulation enables the use of kernel methods. ✔
   (d) The dual formulation can be adapted to the soft margin case.

   Explanation: In the dual formulation, the optimization problem is framed in terms of dot products, so we can directly use a kernel function and solve the optimization in the transformed space.

2. Which of the following is *not* true of the backpropagation algorithm?

   (a) Backpropagation relies on the chain rule for differentiation.
   (b) The gradient for weights in the initial layers is likely to be smaller than those in later layers.
   (c) Backpropagation runs once per minibatch in stochastic gradient descent.✔
   (d) Backpropagation calculations can be speeded up through parallelization.

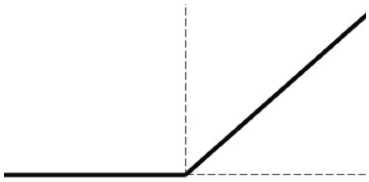
   Explanation: The weights are updated once per minibatch, based on the mean of the gradients in the minibatch. To compute the mean of the gradients, we need to compute the gradient for each input, so backpropagation runs once per input.

3. We feed the output a linear function $z = -3 + 7x$ to a sigmoid function. At what value of x is the centre of the step of the sigmoid?

   (a) 3
   (b) 7
   (c) 3/7 ✔
   (d) 7/3

   Explanation: The step happens where $z = 0$. Set $-3 + 7x = 0$ and solve for x.

4. A rectified linear unit, or ReLU, applies an activation function that converts all negative outputs to zero. Here is a picture of the output of a ReLU as a function of its input.



Let z denote the linear output of the node and a the output of the ReLU activation. Which of the following describe the relationship between a and z?

(a) $a = \max(0, z)$ ✔

(b) $a = \min(0, z)$

(c) $a = |z|$

(d) $a = \alpha z + (1 - \alpha)(1 - z)$

Explanation: For $z < 0$, $0 = \max(0, z)$, for $z > 0$, $z = \max(0, z)$
