Congratulations! You passed!

Grade received 87.50% To pass 80% or higher

Go to next item

0.75 / 1 point

1.	Neural networks are rarely optimised using batch gradient descent (BGD) and instead we use mini-batches from the dataset to train on (often implemented under the name Stochastic Gradient Descent, or SGD). What are the advantages of doing this? Select all that apply.
	☐ SGD gives a better approximation to the true gradient.
	SGD is able to jump out of local minima that would otherwise trap BGD.
	Using BGD is very slow for large datasets, as we need to iterate over the entire dataset before making a single parameter update.
	☐ SGD typically converges in fewer iterations.

2. The binary cross-entropy loss function is a popular choice in classification problems. Let n be the number of training examples, y_i the ground truth label $(y_i \in \{0,1\})$ for the i-th example, and \hat{y}_i the model prediction for the i-th example $(\hat{y}_i \in [0,1])$, i.e. the model probability that the correct class is $y_i = 1$.

1 / 1 point

How is the binary cross-entropy computed for a single instance?

$$\bigcirc (y_i - \hat{y}_i)^2$$

You didn't select all the correct answers

$$\bigcirc |y_i - \hat{y}_i|$$

$$\bigcirc -(y_i \log(1-\hat{y}_i) + (1-y_i) \log(\hat{y}_i))$$

⊘ Correct

Correct, well done!