

# Quiz 1



\* Indicates required question

Email \*

Your email address

the weights may be reduced to zero here \*

1 point

- ☐ L1 and L2
- ☐ L1
- ☐ None of the Above
- ☐ L2

If we want to create an optimal set of weights, we choose to minimize this loss function concerning  $w$  over the entire training data set. \*

1 point

- ☐ True
- ☐ False

What is weight decay? \*

1 point

Your answer

Bagging is an ensemble technique that: \*

1 point

- ☐ Uses a committee of experts to make predictions
- ☐ Trains multiple models on different subsets of the data
- ☐ Combines predictions using a weighted average
- ☐ Constructs an ensemble by iteratively updating weights

Which of these techniques are useful for reducing variance (reducing overfitting)? \*

1 point

- ☐ Dropout
- ☐ L2 regularization
- ☐ Data augmentation

What happens when you increase the regularization hyperparameter  $\lambda$ ? \*

1 point

Your answer

Which of the following functions can be used as an activation function in the output layer if we wish to predict the probabilities of  $n$  classes ( $p_1, p_2, \dots, p_n$ ) such that  $\sum p_i = 1$ ? \*

1 point

- ☐ Tanh
- ☐ ReLu
- ☐ Softmax
- ☐ Sigmoid

Why is the vanishing gradient a problem? \*

1 point

- ☐ The gradient is calculated multiplying two numbers between 0 and 1
- ☐ With backprop, the gradient becomes smaller as it works back through the net
- ☐ All of above.
- ☐ Training is quick if the gradient is large and slow if its small

Which of the following would have a constant input in each epoch of training a Deep Learning model? \*

1 point

- ☐ Weight between input and hidden layer
- ☐ Biases of all hidden layer neurons
- ☐ Weight between hidden and output layer
- ☐ Activation function of output layer

In which of the following applications can we use deep learning to solve the problem? \*

1 point

- ☐ All of the above
- ☐ Protein structure prediction
- ☐ Detection of exotic particles
- ☐ Prediction of chemical reactions

Which of the following is/are Limitations of deep learning? \*

1 point

- ☐ None of the above
- ☐ Obtain huge training datasets
- ☐ Data labeling
- ☐ Both A and B

The purpose of using ensemble learning is to: \*

1 point

- ☐ Increase training time and complexity
- ☐ Eliminate the need for labeled data
- ☐ Reduce overfitting and improve generalization
- ☐ Decrease the number of models required

is a kind of function that tells you how good the prediction of a network is. cross-entropy \*

1 point

- ☐ weight decay
- ☐ Loss function
- ☐ Softmax
- ☐ Softmax Loss

What is TRUE about the functions of a Multi Layer Perceptron? \*

1 point

- ☐ It predicts which group a given set of inputs falls into.
- ☐ It generates a score that determines the confidence level of the prediction.
- ☐ All of above.
- ☐ The first neural nets that were born out of the need to address the inaccuracy of an early classifier, the perceptron.

Select the reason(s) for using a Deep Neural Network \*

1 point

- ☐ Some patterns are very complex and can't be deciphered precisely by alternate means
- ☐ We finally have the technology - GPUs - to accelerate the training process by several folds of magnitude
- ☐ All of the above
- ☐ Deep Nets are great at recognizing patterns and using them as building blocks in deciphering inputs