Congratulations! You passed!
TO PASS 80% or higher

Keep Learning

GRADE
100%

Module 3 Quiz

LATEST SUBMISSION GRADE
100%

1. Select all correct statements: **1 / 1 point**

- [x] Deep Neural Networks are obtained when there are more than two hidden layers.

Correct
Correct!

- [x] Logistic regression can be viewed as a Neural Network with just one sigmoid neuron.

Correct
Correct!

- [ ] A Deep Linear Regression can be obtained if we put "linear neurons" in a hierarchical structure with at least three hidden layers.

- [x] Linear Regression can be viewed as a Neural Network with just one "linear neuron" (a node with a linear activation function).

Correct
Correct!

2. Select all correct statements: **1 / 1 point**

- [ ] The fastest way to find the optimal learning rate for a ML algorithm is to simply add it as one more model parameter and optimize over it, along with other model parameters, in the process of minimization of the train error.

- [x] Making the learning rate variable (larger initially in the training, and smaller as the training progresses) may accelerate ML algorithms.

Correct
Correct!

- [x] A good choice of the learning rate is important: if the learning rate is too small, it takes long for the algorithm to converge, but it if it too high, the algorithm may diverge.

Correct
Correct!

- [ ] Gradient Descent always leads to a unique solution starting from any initial point, no matter what the objective function is.

- [x] Gradient Descent has one free parameter called the learning rate.

Correct
Correct!

3. Select all correct statements: **1 / 1 point**

- [x] Stochastic Gradient Descent attempts at a direct minimization of the generalization error, by producing samples from a data generating distribution in the form of mini-batches.

Correct
Correct!

- [ ] Stochastic Gradient Descent is a practical version of Gradient Descent, named so in recognition of the fact that numerical algorithms often have some numerical noise due to round-up errors etc., so that outputs of Gradient Descent would always be somewhat random.

- [ ] The on-line SGD typically converges much faster than the mini-batch SGD, because in this case there is only one term to evaluate in the loss function.

4. In the last relation in the video on DataFlow and TensorFlow, what is the origin of the factor 2? **1 / 1 point**

Select all correct answers:

- [x] It arises because node `n4` multiplies `x` by itself. Differentiation with respect to the first `x` gives the second `x`, but we can also differentiate the second `x`, which will produce the first `x`.

  ✓ **Correct**
  Correct!

- [ ] The factor 2 arises because there are two variables in the problem, `x` and `y`.

- [ ] It arises because our graph contains exactly two hierarchical levels not counting the last output level. By convention, all derivatives in TensorFlow are multiplied by the depth of the tree, so that this parameter would be easy to extract from the final result.

- [x] This is because the derivative by x of x squared equals 2x.

  ✓ **Correct**
  Correct!

5. Select all correct statements: **1 / 1 point**

- ( ) The Backpropagation algorithm recursively computes first and second derivatives of a train error with respect to all weights of a Neural Network.
- ( ) As backpropagation is a recursive algorithm, it automatically finds a global minimum of a train error, and escapes any potential local minima.
- (•) The Backpropagation algorithm for Neural Networks amounts to Gradient Descent applied to a train error, with a reverse-mode autodiff for a recursive calculation of all derivatives.

  ✓ **Correct**
  Correct!