Suppose you are learning a neural network with momentum. It is given that the learning rate is 0.6, the momentum factor is 0.4, the weight gradient at the ith iteration is 0.56, and the weight increment at the (i-1)th iteration is 0.49. Calculate the weight increment at the ith iteration. Note: Weight at ith iteration (after update) = weight at (i-1)th iteration - weight increment

Answer:

0.14 ✖

The correct answer is: 0.532

Which of the following methods can be used to prevent overfitting? Select one or more:

- [x] a. Regularization ✔
- [x] b. Early Stopping ✔
- [ ] c. Increased Feature Dimensionality
- [x] d. Increased Training Data ✔
- [x] e. Reduced Model Complexity ✔

Suppose you have an input volume of dimension 64x64x16. How many parameters would a single 1x1 convolutional filter have (including the bias)?

Answer:

17 ✔

In a single hidden-layer fully-connected neural network, the input is a vector of dimensionality 7, the number of hidden units is 1, and the number of output neurons is 18. Sigmoid activation functions are used throughout the network. You are supposed to randomly initialize the network weights at the beginning of the training process. The input has an expected distribution of zero mean and unit variance. What should be the standard deviation of random uniform distribution from which the weights tying the input to the hidden units are drawn to prevent the signal levels pushing towards the saturation bands of the activation functions?

Answer:

0.86602540378 ✖

The correct answer is: 0.378

The following equation captures the computation in a ResNet block. What goes into the two blanks above?

$$a^{[l+2]} = g(W^{[l+2]}g(W^{[l+1]}a^{[l]} + b^{[l+1]}) + b^{l+2} + \_\_\_\_\_) + \_\_\_\_\_$$

a. $a^{[l]}$ and 0, respectively ✔

b. 0 and $a^{[l]}$, respectively

c. $z^{[l]}$ and $a^{[l]}$, respectively

d. 0 and $z^{[l+1]}$, respectively

Which ones of the following statements on Residual Networks are true? (Check all that apply.)

- [ ] a. The skip-connections compute a complex non-linear function of the input to pass to a deeper layer in the network.
- [ ] b. A ResNet with L layers would have on the order of L^2 skip connections in total.
- [x] c. Using a skip-connection helps the gradient to backpropagate and thus helps you to train deeper networks
- [x] d. The skip-connection makes it easy for the network to learn an identity mapping between the input and the output within the ResNet block.