**Lecture 14 poll**

**Slide 7: RNNs and MLP**

**Some prediction and classification problems that require very large MLPs and a large amount of training data can be solved using small recurrent nets that only require small amounts of training data**

* True
* False

Some problems that require large, complicated convolutional neural nets and large amounts of training data could also be solved using much smaller RNNs that only require small amounts of training data

* True
* False

**Slide 43: Stability and memory**

**Select all that are true about how long (how many time steps) an RNN can retain some memory of an input pattern**

* It depends on the weights of the recurrent layers
* It depends on the bias of the recurrent layers
* It depends on the activation function used in the recurrent layers
* It depends on the actual input being “remembered”

**Select all that are true about *what* an RNN remembers about an input pattern**

* It depends on the weights of the recurrent layers
* It depends on the bias of the recurrent layers
* It depends on the activation function used in the recurrent layers
* It depends on the actual input being “remembered”

**Slide 66: Vanishing gradient**

**Select all that are true**

* The derivatives for most parameters will become vanishingly small as we backpropagate the loss gradient through deep networks
* The derivatives for a small number of parameters will blow up and become large and unstable as we propagate the los gradient through deep networks
* The derivatives would be more stable if the recurrent weight matrices had singular values equal to 1
* The derivatives would be more stable if the recurrent activations were identity transforms (with identity Jacobian matrices)

**Select all that are true of recurrent networks**

* The memory of the recurrent layer is limited because the recurrent weight matrices are not unitary (with all eigen values equal to 1)
* The memory is also limited by nonlinear activation functions
* The memory would be more stable if the recurrent weight matrix were an identity matrix (i.e. a diagonal matrix with diagonal values equal to “1”)
* The memory would be more stable if the recurrent activations were identity transforms (which are linear and do not scale up or shrink the output)

**Slide 74: LSTMs**

**Select all that are true about LSTMs**

* LSTMs “stabilize” the memory by eliminating the problematic recurrent weights and activations
* They update the memory based on patterns detected in the input and the current context of what they already remember
* In the absence of external cues, they can “remember” a pattern forever
* LSTM are suited to building pattern analyzers requiring long-term memory, e.g. code parsers that can verify if an opened brace has been properly closed