

CS344 Exercise 10

Task 1: Recurrent neural networks

Recurrent neural networks are most commonly used to analyze image data.

TRUE

FALSE

A word embedding is a way to represent words, not as a sequence of characters, but as a vector of numbers such that the representation captures some of the semantic meaning of the words.

TRUE

FALSE

During forward propagation of sequential data through an RNN layer, the output from earlier elements in the sequence have no influence on the output from later elements in the sequence.

TRUE

FALSE

Recurrent neural networks consist only of RNN layers.

TRUE

FALSE

Like max pooling layers, RNN layers normally do not have trainable parameters.

TRUE

FALSE

The hyperbolic tangent function is commonly used as the activation function in RNN layers.

TRUE

FALSE

For some problems, RNNs can be used to predict a single value. For other problems, RNNs can be used to predict a sequence of values.

TRUE

FALSE

Suppose you have an RNN being used for sentiment classification of restaurant reviews. Words are embedded in 25-dimensional space and the first layer of the network is an RNN layer with 16 units. How many total parameters does the first layer of the network have?

Task 2: Architectures for different problems

Suppose you have an RNN with two layers. The first layer is an RNN layer with 16 units and the second (output) layer is a fully-connected (dense) layer. Words are embedded in 25-dimensional space. Fill in the table below describing the network for different problems. Feel free to consult this [helpful site](#) for guidance.

Problem	Activation function for first layer	Total number parameters in first layer	Number of units in output layer	Activation function for output layer	Total number parameters in output layer	Loss function
Sentiment classification with 2 classes			1			
Sentiment classification with 5 classes					85	
Labeling each word one of 2 labels	tanh					
Labeling each word one of 8 labels				softmax		

Task 3: Coding with recurrent neural networks

Download the Jupyter Notebook for Exercise 10 from the course website. Open the Notebook in your web browser and work through it. As you work through the Notebook, answer the following questions.

On what part of speech does the PoS model perform best? On what part of speech does the PoS model perform worst?

What are most of the parameters in the model being used to learn, i.e., what layer has the most parameters that are being trained?

What is the F1 score of the model on the validation data?

How many total parameters does your model have?

What's an example of a "word" generated by the above model that is not actually a word in the English language?

Which of the above models generated text, at least in part, in a language other than English?

CS344 Exercise 10 Final Page

In the *TIME* column, please estimate the time you spent on this exercise. Please try to be as accurate as possible; this information will help us to design future exercises.

PART	TIME
Exercise	