

#### Lecture 11. Recurrent Neural

<u>Course</u> > <u>Unit 3 Neural networks (2.5 weeks)</u> > <u>Networks 2</u>

3. Markov Models to Feedforward

**Neural Nets** 

### **Audit Access Expires May 11, 2020**

You lose all access to this course, including your progress, on May 11, 2020. Upgrade by Mar 25, 2020 to get unlimited access to the course as long as it exists on the site. **Upgrade now** 

# 3. Markov Models to Feedforward Neural Nets Feature Based Markov Models and Temporal/Sequence Problems



 ▶ 0:00 / 0:00

 1.50x

#### Video

Download video file

## **Transcripts**

<u>Download SubRip (.srt) file</u> <u>Download Text (.txt) file</u>

## **Markov Transitions**

2/2 points (graded)

Suppose we represent a Markov model as a feedforward neural network, as described in the lecture. Given a word, let the probability that word j occurs next be  $p_j$ . Which of the condition(s) below must hold true? Let K be the set of words. (Choose all that apply.)



 $lacksquare p_k$  is greater than or equal to zero for all  $k \in K$ 

 $lue{p}_k$  is less than 0.5 for all  $k\in K$ 



How do we satisfy the conditions you marked above? (Choose all that apply.)

take the softmax activation of the outputs

add a bias to the outputs

apply any nonlinear transformation to the inputs



#### **Solution:**

Since it is a probability, it cannot be negative. In addition, as the  $p_k$  represent a probability distribution over the choice of the next word, they must add to 1. As described in the lecture video, a softmax activation forces the probabilities to be

non-negative and sum to 1. Adding a bias and applying a nonlinear transformation don't have anything to do with those two conditions.



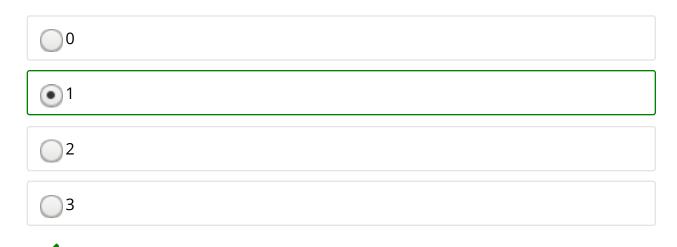
You have used 1 of 2 attempts

**1** Answers are displayed within the problem

### Markov As Feedforward

1/1 point (graded)

When representing a first-order Markov model as a feedforward network, what is the number of non-zero values in a single input vector?



#### **Solution:**

The words are one-hot encoded, so each input word would activate one unique node on the input layer.



You have used 1 of 2 attempts

**1** Answers are displayed within the problem

## Markov vs Feedforward

3/3 points (graded)

What are some advantages of the feedforward NN as described in the lecture versus Markov models? (Choose all that apply.)

They contain a fewer number of parameters

We can easily control the complexity of feedforward NN by introducing hidden layers

They are able to encode more complex transition probabilities than Markov Models.



Suppose you have a word vocabulary of size 10 (including <beg> and <end>), and you were using a trigram language model to predict the next word.

How many parameters would you need for a Markov Model?





How many parameters would you need for a feedforward neural network that contained biases and no hidden units?

190	
195	
200	
<b>2</b> 10	
<b>✓</b>	
Solution:	
	odel would have 100 choices for the previous two words, and 10 choices word, leading to a size of 1000. A feedforward neural network would
	at layer of size 20 and an output layer of size 10, leading to a weight e 200. We add 10 parameters for the bias vector.
Matrix of siz  As demonsti we can add l  architecture encoded in a	e 200. We add 10 parameters for the bias vector.  rated in the second exercise, NNs contain fewer parameters. In addition hidden layers to NNs, showing that they have a more flexible . However, any information encoded in a neural network could also be
As demonstr we can add l architecture encoded in a	e 200. We add 10 parameters for the bias vector.  rated in the second exercise, NNs contain fewer parameters. In addition hidden layers to NNs, showing that they have a more flexible. However, any information encoded in a neural network could also be a very large transition probability matrix, i.e. a Markov Model. Therefore
As demonstrated we can add larchitecture encoded in a the essentia	e 200. We add 10 parameters for the bias vector.  rated in the second exercise, NNs contain fewer parameters. In addition hidden layers to NNs, showing that they have a more flexible. However, any information encoded in a neural network could also be a very large transition probability matrix, i.e. a Markov Model. Therefore I information is the same.
As demonstrated we can add larchitecture encoded in a the essentia	rated in the second exercise, NNs contain fewer parameters. In additional nidden layers to NNs, showing that they have a more flexible. However, any information encoded in a neural network could also be a very large transition probability matrix, i.e. a Markov Model. Therefore I information is the same.  You have used 2 of 2 attempts

3. Markov Models to Feedforward Neural Nets | L... https://courses.edx.org/courses/course-v1:MITx+...

Show all posts

[staff] Markov vs Feedforward problem - terminology consistency
You've called the "bias" in the problem what Professor refers to as the "offset parameter" in t...

[staff] Markov vs Feedforward - NN & Markov input size
Why would we encode the <beg> in both of the words, that constitute the input? It can be onl...

Learn About Verified Certificates

© All Rights Reserved

6 of 6