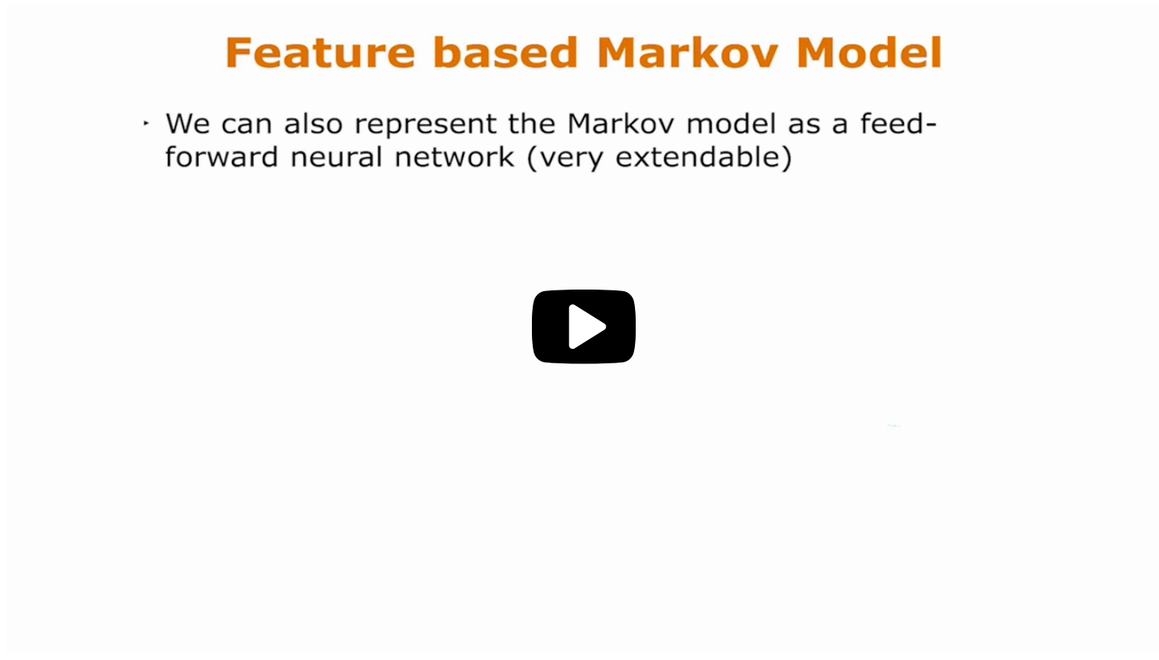
2. Markov Models to Feedforward Neural Nets
Feature Based Markov Models and Temporal/Sequence Problems

Start of transcript. Skip to the end.



0:00 / 0:00 Speed 1.50x

Let's now turn our first order Markov model into a feed-forward neural network model.
To this end, we define an input to the feed-forward neural network.
It is a one hat vector corresponding to the previous word.
And we introduce an input unit for each of the possible previous words.

Video
Download video file

Transcripts
Download SubRip (.srt) file
Download Text (.txt) file

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.)

- [x] $\sum_{k=1} p_k$ ✔
- [x] p_k is greater than or equal to zero for all k ✔
- [ ] p_k is less than 0.5 for all k

✔

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

- [x] take the softmax activation of the outputs ✔
- [ ] add a bias to the outputs
- [ ] apply a 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.

Submit You have used 1 of 2 attempts

Answers are displayed within the problem

Markov As Feedforward

1/1 point (graded)

When representing a Markov model as a feedforward network, how many input nodes have a nonzero value for a given prediction?

- 0
- 1 ✔
- 2
- 3

Solution:

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

Submit You have used 2 of 2 attempts

Answers are displayed within the problem

Markov vs Feedforward

3/3 points (graded)

What are some advantages of feedforward NN versus Markov models? (Choose all that apply.)

- [x] They contain a fewer number of parameters ✔
- [x] 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?

- 1100
- 1001
- 1110 ✔
- 1000 ✔

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

- 190
- 195
- 200 ✔
- 210 ✔

Solution:

A Markov model would have 100 choices for the previous two words, and 10 choices for the next word, leading to a size of 1000. A feedforward neural network would have an input layer of size 20 and an output layer of size 10, leading to a weight matrix of size 200. We add 10 parameters for the bias vector.

As demonstrated in the second exercise, NNs contain fewer parameters. In addition, we can add hidden layers to NNs, showing that they have a more flexible architecture. However,any information encoded in a neural network could also be encoded in a very large transition probability matrix, i.e. a Markov Model. Therefore, the essential information is the same.

Submit You have used 2 of 2 attempts

Answers are displayed within the problem

Discussion

Hide Discussion

Topic: Unit 3 Neural networks (2.5 weeks):Lecture 11. Recurrent Neural Networks 2 / 2. Markov Models to Feedforward Neural Nets

Add a Post

All Posts

question 2

question posted a day ago by **BOBQian**

I don't understand the question after reading multiple times, can anyone help?When representing a Markov model as a feedforward network, how many input nodes have a nonzero value for a given prediction?

This post is visible to everyone.

Add a Response

2 responses

disguiser
a day ago

It asks about the number of non-zero values in a single input vector.

Doesn't this depend on the depth of the input (how many input words are used as input)? Shouldn't the question specify the order of the Markov model?

posted a day ago by **stevenaronson**

+1

posted a day ago by **sandipan_dey**

Add a comment

nr7116
about 3 hours ago

The question is answered in the video around 0:30 seconds
however, I don't know what he means by one-hot encoding.

in general one hot encoding is when you encode a categorical attribute using a feature vector of all zeroes for the non satisfied categories and a single 1 to represent the satisfied category. so in this case i'm pretty sure it refers to encoding a word by a using a vector of all zeroes for the whole vocabulary and a single 1 for the word it represents.

posted about 2 hours ago by **synnfusion**

Add a comment

Showing all responses

Add a response:

Preview

Submit

Learn About Verified Certificates