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

and mapping the combination of preceding words

to first the hidden layer and then to the output layer.

So depending on the number of hidden units

that we introduce here, we can increase the complexity

of the information that the neural network

model can make use of from the pair of preceding words,

in order to predict what comes next.

End of transcript. Skip to the start.

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] $\sum_{k \in K} p_k = 1$ ✔
- [x] p_k is greater than or equal to zero for all $k \in K$ ✔
- [ ] 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.)

- [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 1 of 2 attempts

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.)

- [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

Show Discussion

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