Lecture 12 Quiz

Quiz, 7 questions

1 point

1.
A Boltzmann Machine is different from a Feed Forward Neural Network in the sense that :

- [ ] The state of a hidden unit in a Boltzmann Machine is a deterministic function of the inputs and is hard to compute exactly, but in a Neural Net it is easy to compute just by doing a forward pass.
- [x] The state of a hidden unit in a Boltzmann Machine is a random variable, but in a Neural Net it is a deterministic function of the inputs.
- [x] A Boltzmann Machine defines a probability distribution over the data, but a Neural Net defines a deterministic transformation of the data.
- [ ] Boltzmann Machines do not have hidden units but Neural Nets do.

---

1 point

2.
Throughout the lecture, when talking about Boltzmann Machines, why do we talk in terms of computing the **expected** value of $s_i s_j$ and not the value of $s_i s_j$?

- ( ) It is not possible to compute the exact value no matter how much computation time is provided. So all we can do is compute an approximation.
- ( ) It is possible to compute the exact value but it is computationally inefficient.
- (•) It does not make sense to talk in terms of a unique value of $s_i s_j$ because s_i and s_j are random variables and the Boltzmann Machine defines a probability distribution over them.

Lecture 12 Quiz

Quiz, 7 questions

- ( ) The expectation only refers to an average over all training cases.

1 point

3.

When learning an RBM, we decrease the energy of data particles **and** increase the energy of fantasy particles. Brian insists that the latter is not needed. He claims that it is should be sufficient to just decrease the energy of data particles and the energy of all other regions of state space would have increased relatively. This would also save us the trouble of sampling from the model distribution. What is wrong with this intuition ?

- ( ) The sum of all updates must be zero so we need to increase the energy of negative particles to balance things out.
- ( ) Since total energy is constant, some particles must loose energy for others to gain energy.
- ( ) There is nothing wrong with the intuition. This method is an alternative way of learning a Boltzmann Machine.
- (•) The model could decrease the energy of data particles in ways such that the energy of negative particles also gets decreased. If this happens there will be no net learning and energy of all particles will keep going down without bounds.

1 point

4.

Restricted Boltzmann Machines are easier to learn than Boltzmann Machines with arbitrary connectivity. Which of the following is a contributing factor ?

- ( ) The energy of any configuration of an RBM is a linear function of its states. This is not true for a general BM.
- (•) In RBMs, there are no connections among hidden units or among visible units.
- ( ) It is possible to run a persistent Markov chain in RBMs but not in general BMs.
- ( ) RBMs are more powerful models, i.e., they can model more probability distributions than general BMs.

Lecture 12 Quiz

Quiz, 7 questions

1 point

5.

PCD a better algorithm than CD1 when it comes to training a good generative model of the data. This means that samples drawn from a freely running Boltzmann Machine which was trained with PCD (after enough time) are likely to look more realistic than those drawn from the same model trained with CD1. Why does this happen ?

- [x] In PCD, the persistent Markov chain explores different regions of the state space. However, CD1 lets the Markov chain run for only one step. So CD1 cannot explore the space of possibilities much and can miss out on increasing the energy of some states which ought to be improbable. These states might be reached when running the machine for a long time leading to unrealistic samples.
- [ ] In PCD, many Markov chains are used throughout learning, whereas CD1 uses only one. Therefore, samples from PCD are an average of samples from several models. Since model averaging helps, PCD generates better samples.
- [ ] In PCD, only a single Markov chain is used throughout learning, whereas CD1 starts a new one in each update. Therefore, PCD is a more consistent algorithm.
- [ ] In PCD, the persistent Markov chain can remember the state of the positive particles across mini-batches and show them when sampling. However, CD1 resets the Markov chain in each update so it cannot retain information about the data for a long time.

1 point

6.

I'ts time for some math now!

Lecture 12 Quiz

Quiz, 7 questions

In RBMs, the energy of any configuration is a linear function of the state.

$E(\mathbf{v},\mathbf{h}) = -\sum_i a_i v_i - \sum_j b_j h_j - \sum_{i,j} v_i h_j W_{ij}$

and this eventually leads to

$\Delta W_{ij} \propto \langle v_i h_j \rangle_{data} - \langle v_i h_j \rangle_{model}$

If the energy was non-linear, such as

$E(\mathbf{v},\mathbf{h}) = -\sum_i a_i v_i - \sum_j b_j h_j - \sum_{i,j} f(v_i, h_j) W_{ij}$

for some non-linear function f, which of the following would be true.

- ( ) $\Delta W_{ij} \propto \langle v_i h_j \rangle_{data} - \langle v_i h_j \rangle_{model}$
- ( ) $\Delta W_{ij} \propto f(\langle v_i \rangle_{data}, \langle h_j \rangle_{data}) - f(\langle v_i \rangle_{model}, \langle h_j \rangle_{model})$
- (•) $\Delta W_{ij} \propto \langle f(v_i, h_j) \rangle_{data} - \langle f(v_i, h_j) \rangle_{model}$
- ( ) The weight update can no longer be written as a difference of data and model statistics.

1
point

7.
In RBMs, the energy of any configuration is a linear function of the state.

$E(\mathbf{v},\mathbf{h}) = -\sum_i a_i v_i - \sum_j b_j h_j - \sum_{i,j} v_i h_j W_{ij}$

and this eventually leads to

$P(h_j = 1|\mathbf{v}) = \frac{1}{1+\exp(-\sum_i W_{ij} v_i - b_j)}$

If the energy was non-linear, such as

$E(\mathbf{v},\mathbf{h}) = -\sum_i a_i v_i - \sum_j b_j h_j - \sum_{i,j} f(v_i, h_j) W_{ij}$

for some non-linear function f, which of the following would be true.

- ( ) $P(h_j = 1|\mathbf{v}) = \frac{1}{1+\exp(-\sum_i W_{ij} v_i - b_j)}$
- ( ) None of these is correct.
- ( )

$$P(h_j = 1|\mathbf{v}) = \frac{1}{1+\exp(-\sum_i W_{ij}(f(v_i,1)+f(v_i,0))-b_j)}$$

Lecture 12 Quiz

Quiz, 7 questions

$$P(h_j = 1|\mathbf{v}) = \frac{1}{1+\exp(-\sum_i W_{ij}(f(v_i,1)-f(v_i,0))-b_j)}$$

I, **Alan Wright**, understand that submitting work that isn't my own may result in permanent failure of this course or deactivation of my Coursera account.

Learn more about Coursera's Honor Code

Submit Quiz