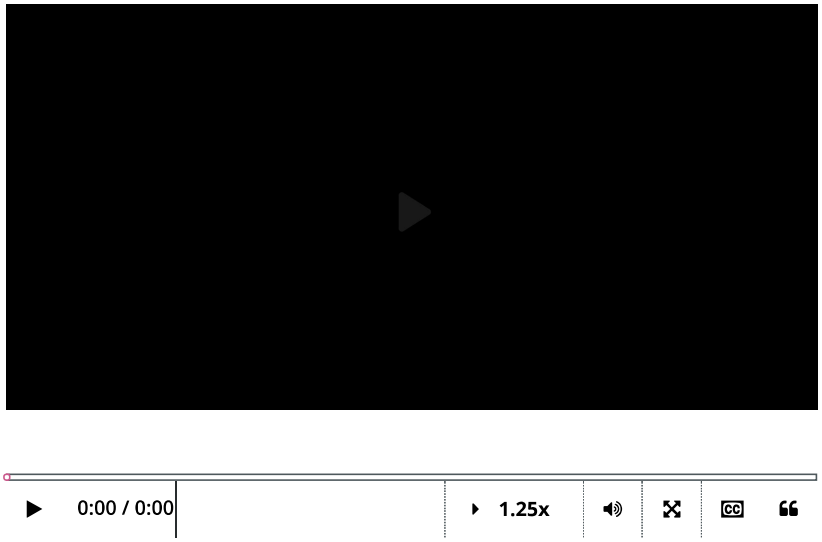
3. Training Models with 1 Hidden Layer

Exercises due Oct 28, 2020 05:29 IST *Completed*

Video note: In the video below at 0:38, Prof Jaakkola mispoke and said “on the left," but you should look at the plot “on the **right**".

Training Models with 1 Hidden Layer, Overcapacity, and Convergence Guarantees



Start of transcript. Skip to the end.

So let's take here a few visual examples,

just to see how the network evolves

as a result of stochastic gradient descent updates.

Here, I have represented an initial network,

where I have the two input coordinates-- x1. x2--

Video
Download video file

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

SGD Convergence guarantees

1/1 point (graded)
Which of the following option(s) is/are true about training neural networks?
(Choose all that apply.)

- [ ] For multi-layer neural networks, stochastic gradient descent (SGD) is guaranteed to reach global optimum
- [x] For multi-layer neural networks, stochastic gradient descent (SGD) is not guaranteed to reach a global optimum
- [ ] Larger models tend to be harder to learn because their units need to be adjusted so that each one of them can individually solve the task
- [x] Larger models tend to be easier to learn because their units need to be adjusted so that they are, collectively sufficient to solve the task
- [ ] Initialization plays no or very little role in finding a good solution during training of neural networks

✔

Solution:

- For multi-layer neural networks the loss function is no longer convex and any stochastic gradient descent (SGD) method is not guaranteed to reach global optimum

- Larger models tend to be easier to learn because their units need to be adjusted so that they are, collectively sufficient to solve the task

Submit You have used 1 of 2 attempts

Answers are displayed within the problem

Discussion

Hide Discussion

Topic: Unit 3 Neural networks (2.5 weeks):Lecture 9. Feedforward Neural Networks, Back Propagation, and Stochastic Gradient Descent (SGD) / 3. Training Models with 1 Hidden Layer

Add a Post

Show all posts by recent activity

A visual proof that neural nets can compute any function (Link) Neural Nets can (at least in theory) compute any function based on the universal approximation theorem. More info and visuals here: ht... Pinned Community TA	3
random initialization Quoting the professor "The randomization inherent in the initialization creates some smoothness. And we end up with a smooth decisio...	1
Marks are not shown properly Hi! I have a 5 of 5 for this lecture, but in the progress section it shows me only 80% of success and all exercises marked as 1/1.... I am a b...	2
Some thoughts and questions!	4
SGD Convergence Guarantees: Double Entendre => ATTN STAFF (A 20 MARKS GAMBLE)!!! Hi, The question has as a last answer option: "Initialization plays no or very little role in finding a good solution during training of neural ...	15
Confused I feel that I don't understand a bit. I think I need to read other sources to get more intuitive and formal understanding on deep network...	3
Neural networks vs linear regression In the previous unit we spent a lot of time learning how increasing the dimension via non-linearities helps solve complicated training mo...	2
Is there some kind of method to choose the architecture of the NN? I understand that we can explode the overcapacity and the random offset initialization in order to find quite good decision boundaries. ...	3

Previous Next

Legal

Terms of Service & Honor Code
Privacy Policy
Accessibility Policy
Trademark Policy
Sitemap

Connect

Blog
Contact Us
Help Center
Media Kit
Donate

/