3. Training Models with 1 Hidden Layer

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

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

? offset parameter 1

Why is it that we randomise the choice of offset parameter the redundant part that was appe...

Learn About Verified Certificates