Questions and answers - Lukas Heinrich Lecture 2

The following questions were submitted through Google Form. Some may have been answered in the Q&A session already. Nevertheless, we request our lecturers to provide written answers here for the benefit of those who could not attend that session. Thank you!

Slide about shallow vs deep NN. The plots showing the respective losses over time for shallow and deep networks: Were the numbers for each network the number of neurons, or the number of parameters (deeper networks have more parameters per neuron, right?)

LH: overall number of parameters

Slide 31. When you talked about why deep architectures learn more efficiently than shallower architectures, you said "we THINK this is why". Can you speak to how much we actually understand about how these networks work / how much interpretability analysis has gone on?

LH: this is still an active research area. There is quite some research into what vision models learn. A interesting start could be to read this: https://distill.pub/2020/circuits/zoom-in/

Slide not specified. Is there a way to encode rotational invariance into CNN? Is this even a good idea?

LH: https://www.reddit.com/r/dataisbeautiful/comments/aydqig/is_it_a_duck_or_a_rabbit_for_google_cloud_vision/

Slide 22. Softmax looks like Boltzman distribution. Is there any physics-analogy interpretation?

LH: Yes, the normalization turns raw energies into normalized densities like the partition

Slide 26. What are the examples of exhaustive search and closed form solutions?

LH: hyperparameter optimization is often done through exhaustive search in a finite set. Polynomial Regression Models can be solved in closed form (see Lectuer 1)

Slide 59. I notice that in some cases such as Unet, pooling layer is applied after every two convolutional layers; and in some other cases such as ResNet, the pooling layer is only applies after the first and last convolutional layers. Could you explain why? and how researchers make decision on adding or not adding pooling layer?

LH: there is no hard and fast rule here

Slide 72. Could you further explain how to generate the right-half of the subplots in the figure 3?

LH: Should have added a reference: the figure is from here: https://arxiv.org/pdf/1502.03044.pdf