

2022-02-07: Questions and answers

Made with a warm hug

CHEERYTEXT3944 JAN 20, 2022 09:24PM

Training

I think it would help if there is an expected runtime for a given task in the lab description, that can be used to tell if you are doing it right

Thank you for the suggestion! It is a bit tricky to do that b/c of the many different hardware configurations that students could have. As a rough indication, our reference implementation takes about 50 seconds per epoch on Colab (no GPU support).
– CHEERYTEXT3944

Why is accuracy used and not F1? I thought F1 was better for unbalanced classes and I am guessing some parts of speech are way more popular than others.

Good point! For an actual error analysis, you would want to focus on individual tags and use precision/recall/F1. Here the focus is not so much on the evaluation itself, but on the model; computing the accuracy is basically only used to test your code.
– CHEERYTEXT3944

Is this a computationally heavy lab?

Compared to L2, the lab is rather lightweight. On Colab, our reference implementation takes about 50 seconds per epoch without GPU support. – CHEERYTEXT3944

Will the winners of the chocolate challenge be draw at the end of the course after final submission?

No, we will announce them in next week's session! 😊 (You can still get credit for the lab even if you do not participate in the challenge.) – CHEERYTEXT3944

What's the purpose of the ReLU in the feed forward network?
And the purpose of the feed forward network in general

*The idea behind the FFNN is to create a useful representation for part-of-speech tagging that goes beyond the concatenation of embeddings that we create in the first part of the model. This is done in a standard way: a two-layer neural network with an intermediate, non-linear transfer function. Here, ReLU is the go-to choice. (There are reasons for that:
<https://stats.stackexchange.com/questions/126238/what-are-the-advantages-of-relu-over-sigmoid-function-in-deep-neural-networks>)* – CHEERYTEXT3944

It seems reasonable to squeeze 150->50 in the linear layer...

But squeezing 160->50 seems much stranger, what should you consider while doing that squeeze

I guess this is largely an empirical question. The basic intuition is that smaller dimensions force the model to generalise more, but there is no formula that tells us how much we can squeeze and still achieve a certain performance. – CHEERYTEXT3944

RNN sequence labeling

Can RNNs be used for sequence labeling?

Absolutely; see Lecture 3.2. – CHEERYTEXT3944
