Initial Thesis Ponderings

Rasmus Diederichsen

2018-05-11

Who is he?

Figure 1:rasmus

What does he like?

- Frontend vs. Backend: **backend**
- Vim vs. Emacs: **vim**
- Tabs vs. Spaces: **spaces** (fight me)
- Dynamic vs. Static: **undecided**
- Clean code
- Documentation (<3 sphinx)
- Automation

What doesn't he like?

- Web programming

Project goals

Create an open source live monitoring framework for deep neural networks.

Engineering perspective

In an ideal world, the following requirements would be met:

- Be agnostic to Pytorch and TF/Keras
- Provide a platform for easy experimentation with arbitrary diagnostics
- Enable live monitoring/insights as well as explanation for model predictions
- Build on top of Tensorboard
- Bring to life in Peltarion platform

Risk:

What has he done?

So far, I've

- skimmed a few explanation papers (LIME, influence functions)
- skimmed many more visualization publications (grad-cam, heatmapping)
- passively absorbed some ideas from Tishby's information bottleneck theory (status unclear)
- Started learning Pytorch by programming myself into a corner
  - Turns out my toy framework wouldn't live up to my plans -> **rebase onto `tensorboard-pytorch`**
- Compiled notes to distill into roadmap

Goals for this visit

- Get to know people and products
- Pick platform brains
  - What visualizations/introspections exist or are planned?
  - (How) can I interface with the backend platform?
  - what would you like from an introspection toolbox?

Idea junkyard

Detecting bad initializations

- Monitor Jacobian determinant and singular values of layer function (w.r.t. prev layer, c.f. Gloro init paper)
- Inject prior knowledge about e.g. image feature extractors (if they don't converge to edges, maybe they're stuck)

Architectural Problems

- Activation eigenspace analysis (cue Justin)
- Monitor sparsity of encodings/detect correlations

Saturation

- Tishby hinted that his info bottleneck theory could help detect layer saturation (where did I read this?)
- Monitor ratio between weights and updates

Learning rate issues