Jost BA — ADAM

Theory

- origin and motivation of adam
- consise and clear formulation of adam and its parameters
- explain backprop?
- motivate techniques to improve learning (dropout, early stopping, regularization, normalization, ...)

Implemenation / Experiments

- start with TF implemenation for standard benchmark problems and dense networks parameter studies for benchmarks (activation functions, network layout, layer types (ResNet?), hyperparameters of adam) — document your procedure / progress
- implement ChemEx
- parameter studies for this hard and unknown problem (activation functions, network layout, layer types (ResNet?), hyperparameters of adam) — document your procedure / progress

Goals

- solid understanding of current state of the art optimization algorithm (adam)
- collection of commonly used techniques to improve learning (dropout, early stopping, regularization, normalization, ...)
- hopfully improve results on ChemEx
- in house comparison data for MNIST or other benchmark problems
- show that adam is very powerfull but hard to tune?
- stable and well documented code

Offtopics

- line search type algorithms in TF with CustomModel and CustomOptimizer
- real line search GD for small problems (compare to adam)
- Augementation of the inputdimension in ResNet

Organization

- (common) GIT?
- regular meetings?
- timeline?
- seminartalk at mid and/or end of timeline (no official defence is required)