STOCHASTIC OPTIMIZATION IN MACHINE LEARNING

Case Studies in Nonlinear Optimization

F. Bauer   S. Chambon   R. Halbig   S. Heidekrüger   J. Heuke

July 5, 2015

Technische Universität München

We're not running out of data anytime soon. It's maybe the only resource that grows exponentially.

Andreas Weigend

INTRODUCTION

- Massive amounts of training data
- Construction of very large models
- How to handle the high memory/computational demands?

- Massive amounts of training data
- Construction of very large models
- How to handle the high memory/computational demands?

Solution: Stochastic Methods
Update on smaller amounts of training data in each iteration!

A STOCHASTIC QUASI-NEWTON METHOD

What is it? Why? Main ideas, high-level pseudo code overview? short bfgs repitition? Extreme Cases (L-BFGS, SGD)

Explain the Dataset quickly. Why is this good for SQN testing? Why is it challenging? (file size etc)

Pretty picures about the behaviour of SQN on HIGGS and comparison with traditional SGD

PROXIMAL SPLITTING METHOD

Outline of Section by Jacob and Fin

- hallo
- du muschi
- [1]

LOGISTIC REGRESSION: AN EXAMPLE

Explain what we want to do, and explain the dataset, and why using both SQN and Prox makes sense
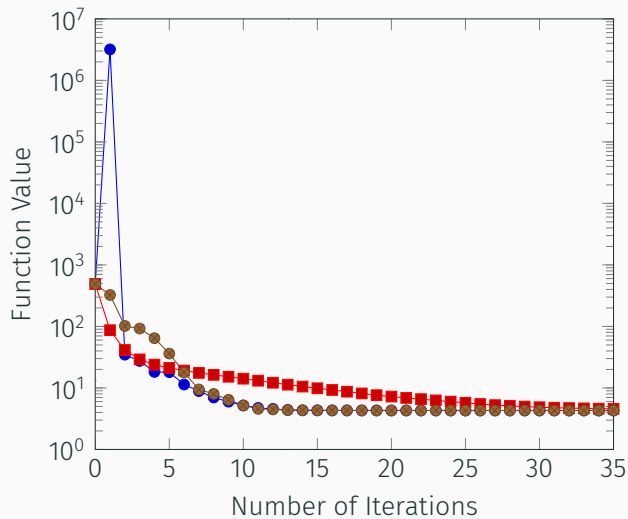
Nice table with SQN, SGD (no reg, L2), (Lasso,) Prox (L1) showing Obj. value in found optimum, CPU time, Iterations, F1 score of prediction model

Use different reg. parameters?? Stop after fixed time? after fixed iters? after insign. improvements

CONCLUSION

hallo

QUESTIONS?

📄 S. Becker and J. Fadili.
A quasi-newton proximal splitting method.
In *Advances in Neural Information Processing Systems*, pages 2618–2626, 2012.