

High-dimensional local optimization in variational inference

Otte Heinävaara

Supervisor: Antti Honkela
Probabilistic Inference and Computational Biology (PROBIC)
Statistical Machine Learning and Bioinformatics
HIIT

August 26, 2015

Variational inference

Bayes' rule

$$- P(Z|E) = \frac{P(Z)}{P(E)} P(E|Z)$$



Bayesian inference

$$- P(Z|E) = \frac{P(Z)}{\sum_i P(E|Z_i)P(Z_i)} P(E|Z)$$



Variational inference

$$- P(Z|E) \approx Q(Z)$$

How does a machine discover a topic?

This study is about the science of food. Do you like pizza or pasta? Is there way to determine how tasty food is without tasting or smelling it?

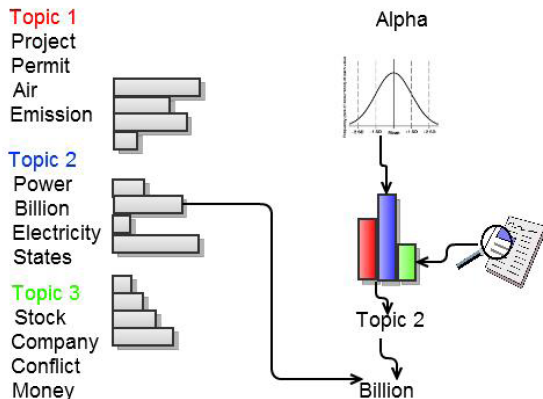


This **study** is about the **science** of **food**. Do you like **pizza** or **pasta**? Is there way to **determine** how **tasty food** is without **tasting** or **smelling** it?

Words related to **food**
science

LDA - Latent Dirichlet Allocation

D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent Dirichlet allocation. *The Journal of Machine Learning Research*, 3:993-1022, 2003.

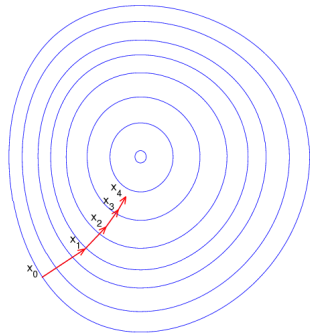
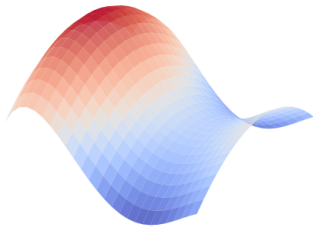


Picture: Amogh Mahapatra, Nisheeth Srivastava and Jaideep Srivastava. Contextual Anomaly Detection in Text Data, *Algorithms 2012*, pages 469-489, figure 3, 2012.

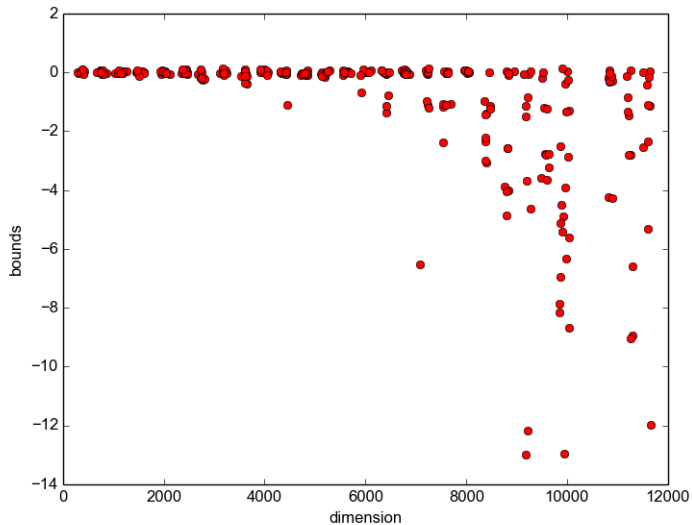
Underlying math (essentially)

1. Use LDA to interpret the topic assignment as a problem in Bayesian inference
2. Variational inference:
 - ▶ Using suitable family of approximations and metric, the problem is turned to optimization in a metric space
 - ▶ Doing suitable estimations and assumptions, problem is moved to R^n and the target function (i.e. the metric) becomes tractable to compute

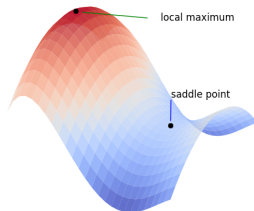
High-dimensional optimization



Effect of dimension



Where do we end up?

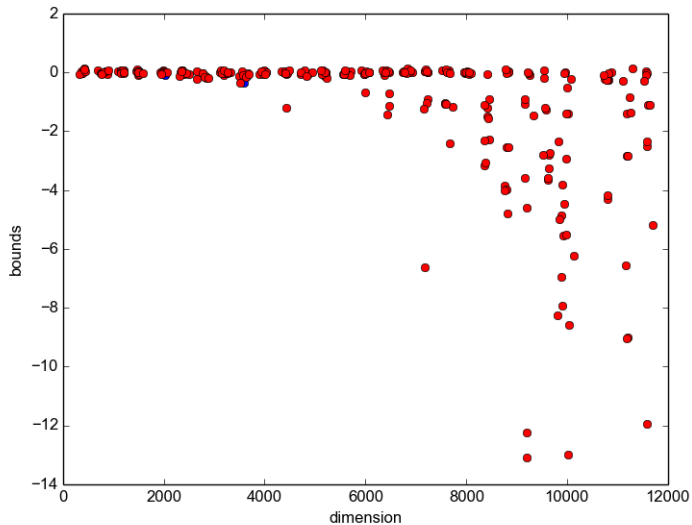


Curse of dimensionality

“...in high dimensions, the chance that all the directions around a critical point lead upward (positive curvature) is exponentially small...” (Dauphin et al. 2014¹)

¹Yann N Dauphin, Razvan Pascanu, Caglar Gulcehre, Kyunghyun Cho, Surya Ganguli, and Yoshua Bengio. Identifying and attacking the saddle point problem in high-dimensional non-convex optimization. *In Advances in Neural Information Processing Systems*, pages 2933-2941, 2014.

Effect of dimension, eigenvalues



What next?

- ▶ More info about the eigenvalues/local behaviour
- ▶ Improving methods

The main things

To sum up:

- ▶ Turning topic assignment to high-dimensional optimization via
 1. building a model
 2. using variational ideas to deal with Bayesian inference
- ▶ In high-dimensional spaces, many simple ideas lose their edges