ATKearney

Big Data Konferenz #3

Bayesian Optimization

Presentation
June 09, 2016

Startplatz, Cologne

Dr. Paul Viefers

Agenda

- **About A.T. Kearney**
- The idea behind Bayesian Optimization
- Earning while learning: A/B testing
- Tuning algorithms with BO
- Practical considerations

About A.T. Kearney

A.T. Kearney is a leading global management consulting firm with offices in more than 40 countries. Since 1926, we have been trusted advisors to the world's foremost organizations. A.T. Kearney is a partner-owned firm, committed to helping clients achieve immediate impact and growing advantage on their most mission-critical issues.

In serving our clients, we combine our global reach with local expertise

Clients	We work with more than two-thirds of the Fortune Global 500, the world's largest companies by revenues, as well as with the most influential governmental and non-profit organizations.
Locations	A.T. Kearney has 61 offices located in major business centers in more than 40 countries.
Team	We are 3,500 people strong worldwide who have broad industry experience and come from leading business schools. We staff client teams with the best skills for each project from across A.T. Kearney.

Americas	Atlanta Bogotá Calgary Chicago Dallas	Detroit Houston Mexico City New York Palo Alto	San Francisco São Paulo Toronto Washington, D.C.
Asia Pacific	Bangkok Beijing Hong Kong Jakarta Kuala Lumpur	Melbourne Mumbai New Delhi Seoul Shanghai	Singapore Sydney Taipei Tokyo
Europe	Amsterdam Berlin Brussels Bucharest Budapest Copenhagen Düsseldorf Frankfurt Helsinki	Istanbul Kiev Lisbon Ljubljana London Madrid Milan Moscow Munich	Oslo Paris Prague Rome Stockholm Stuttgart Vienna Warsaw Zurich
Middle East and Africa	Abu Dhabi Doha	Dubai Johannesburg	Manama Riyadh

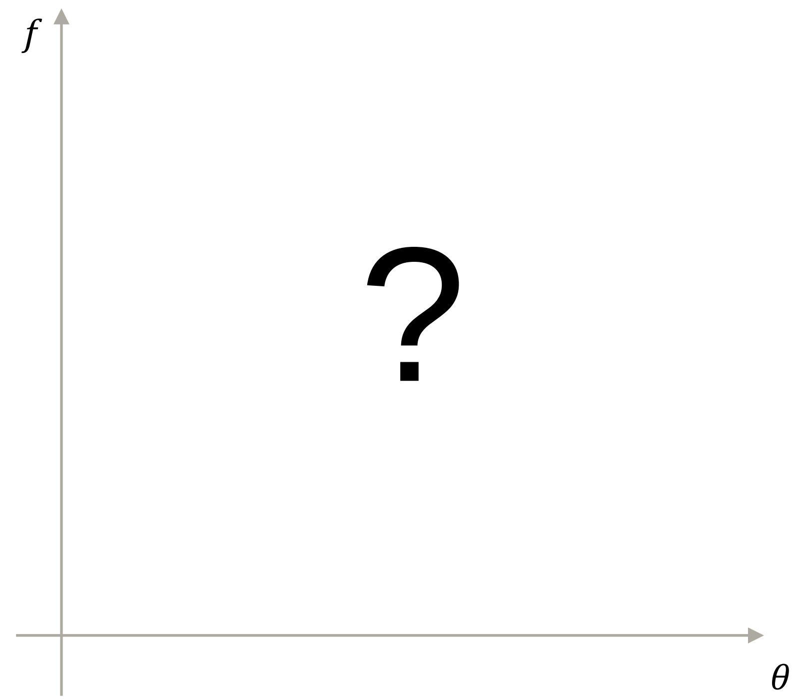
Agenda

- About A.T. Kearney
- **The idea behind Bayesian Optimization**
- Earning while learning: A/B testing
- Tuning algorithms with BO
- Practical considerations

Data Analysts often have to find extreme points of black-box functions

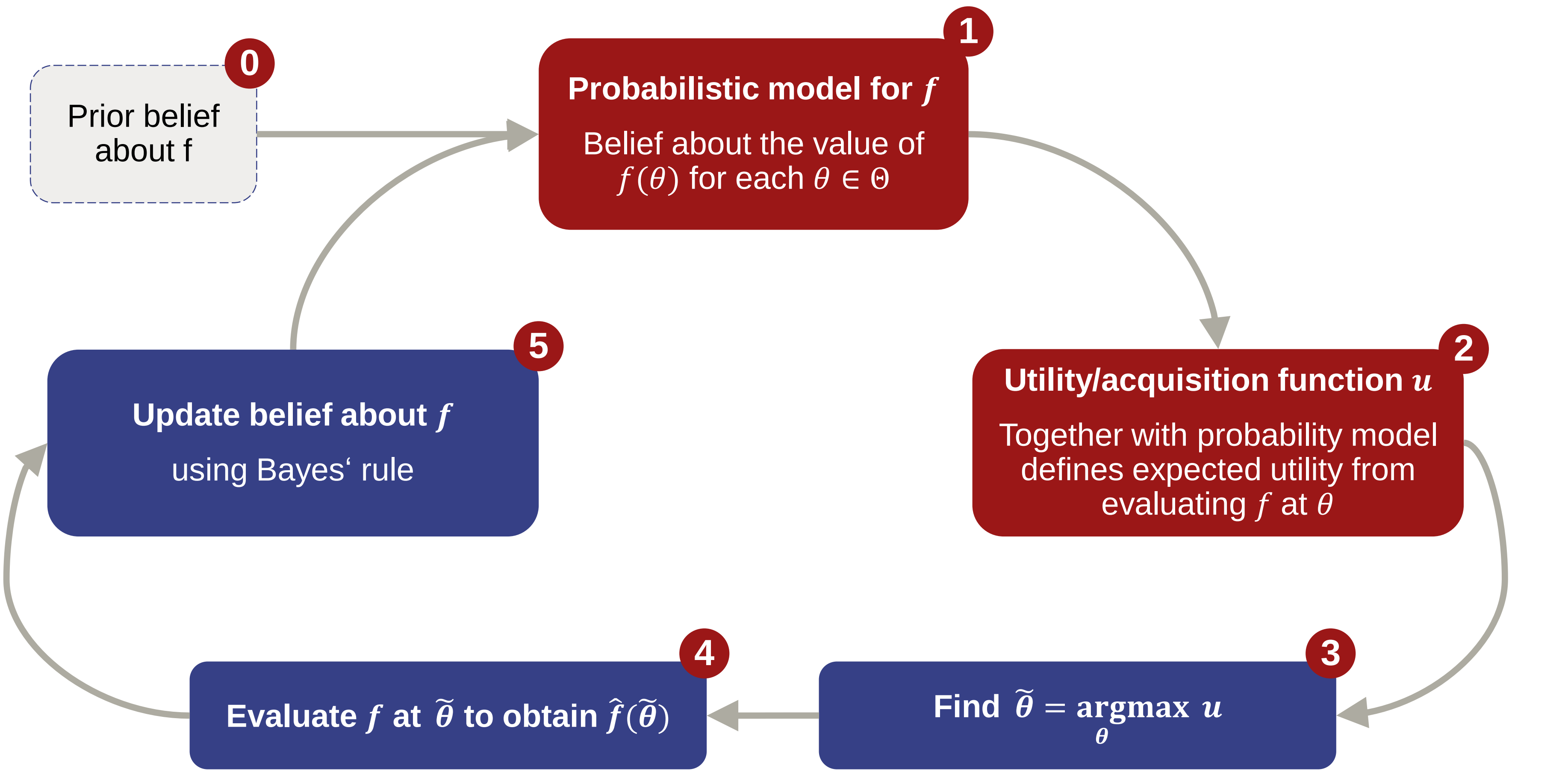
Situation

- Want to maximize/minimize objective function $f\colon \Theta \to \mathbb{R}$
- Domain Θ potentially high-dimensional
- Do no know (functional) form of f
- Do not have gradients available
- Can evaluate f everywhere, but the return value $\hat{f}$ is potentially noisy
- Evaluating f is expensive (time and/or money)

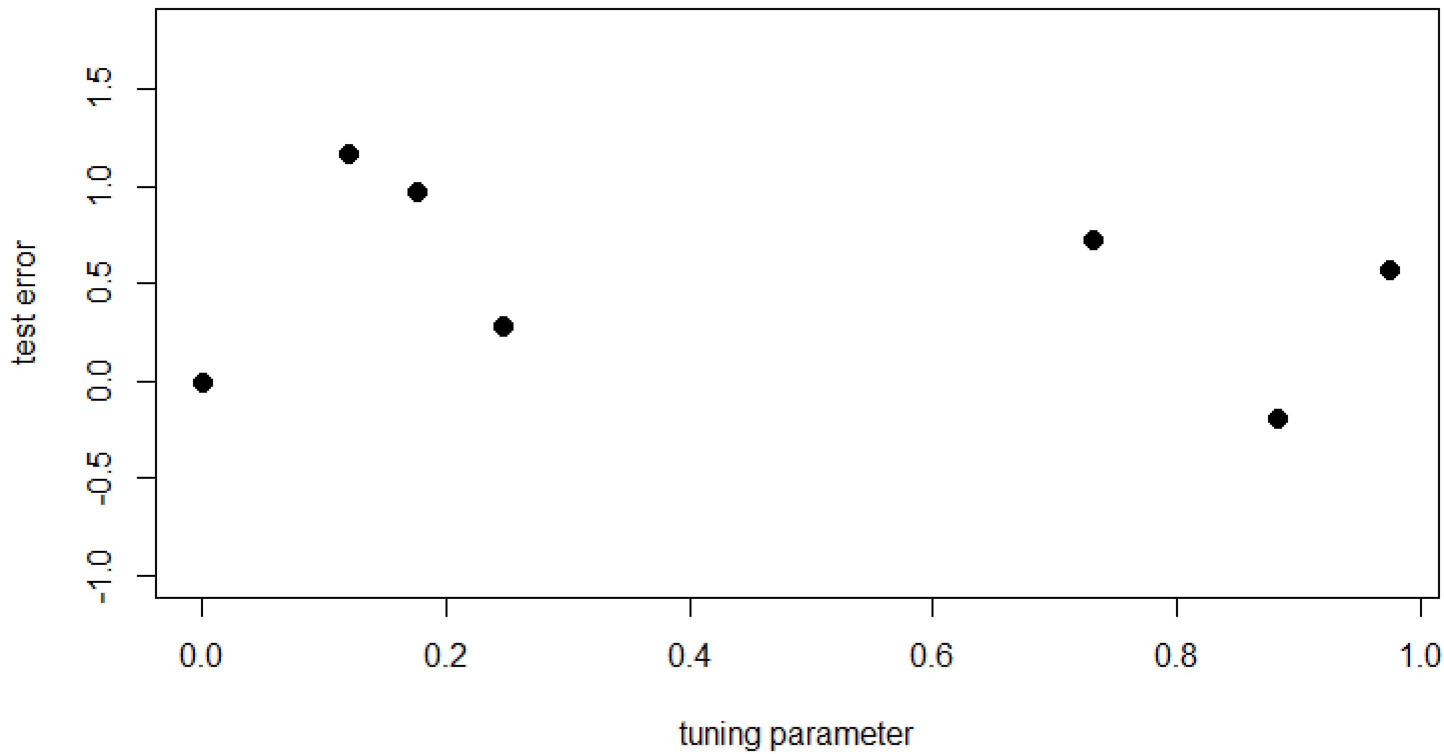


ATKearney

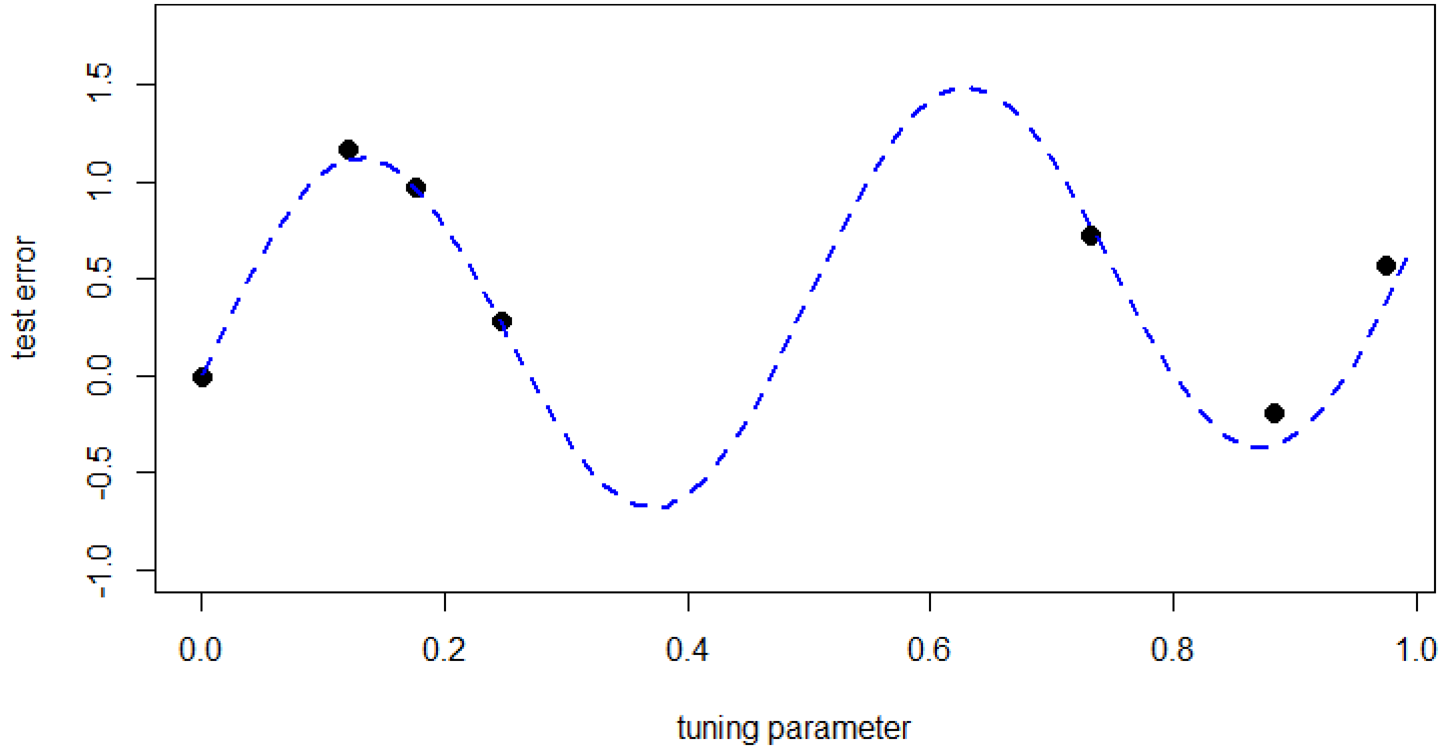
Bayesian Optimization provides a rigorous framework to find extreme points in this setup



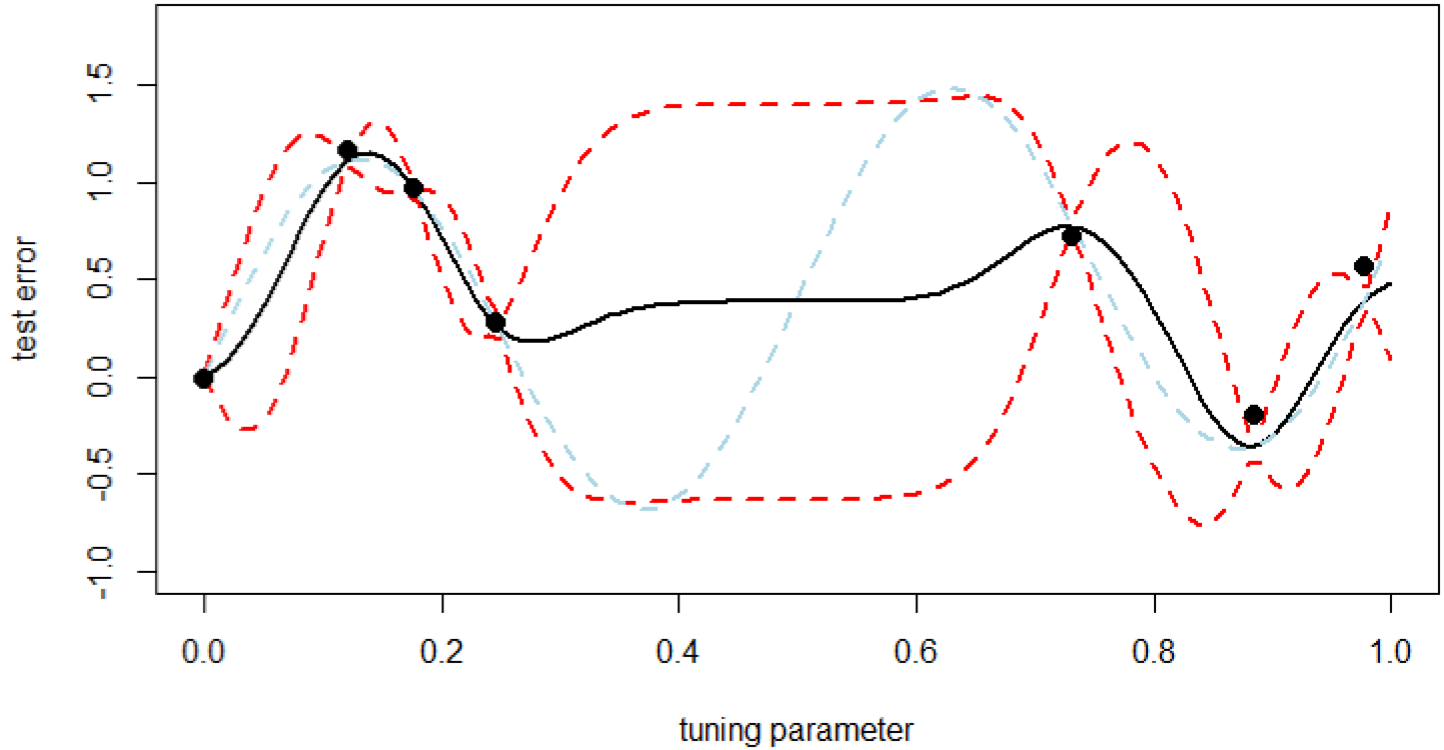
Motivation: Training a statistical model



Motivation: Training a statistical model

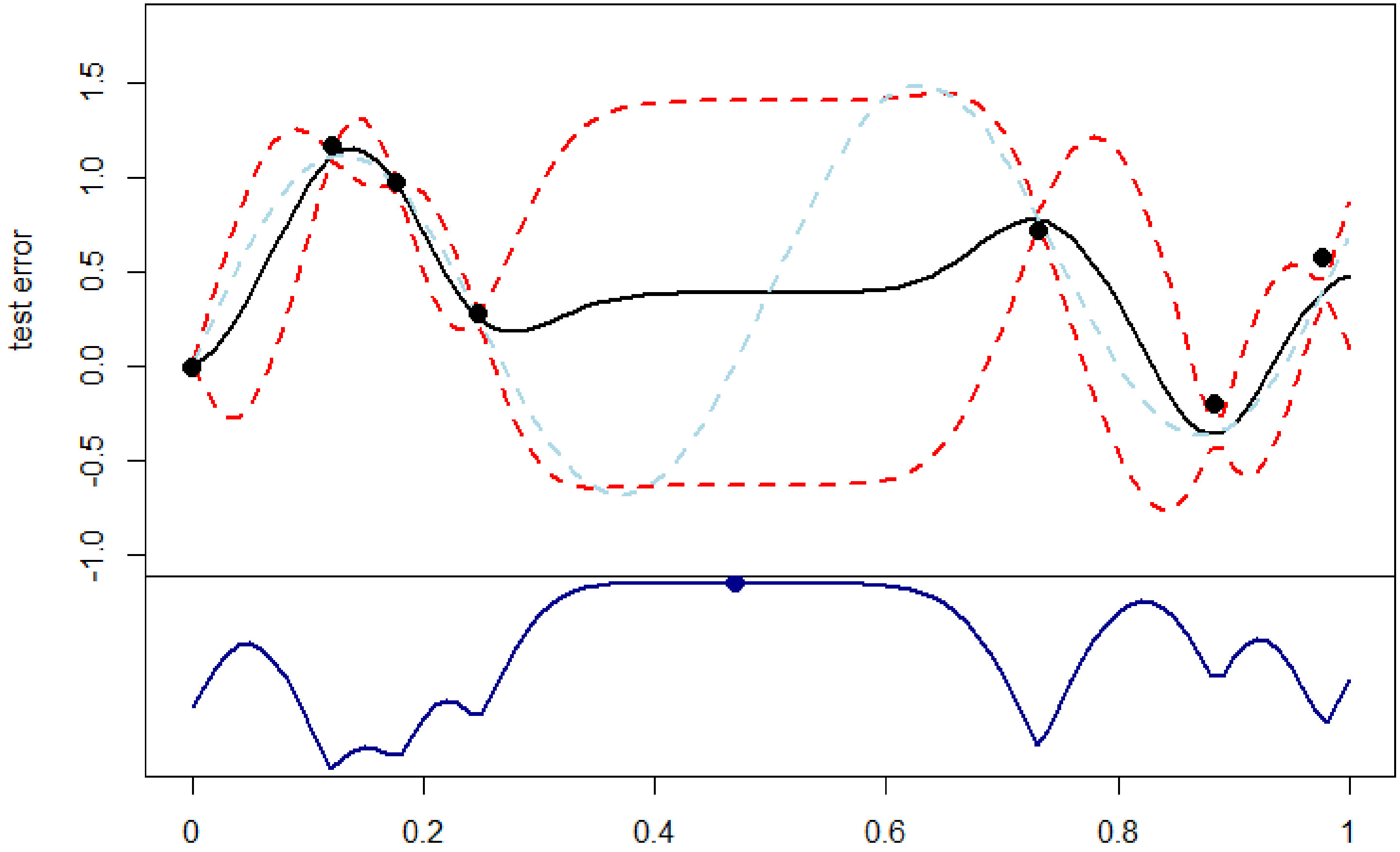


Motivation: Training a statistical model



ATKearney

Motivation: Training a statistical model



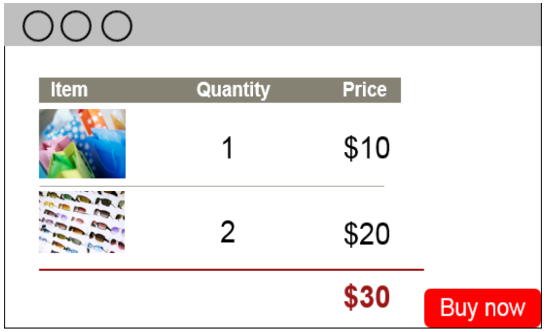
Agenda

- About A.T. Kearney
- The idea behind Bayesian Optimization
- **Earning while learning: A/B testing**
- Tuning algorithms with BO
- Practical considerations

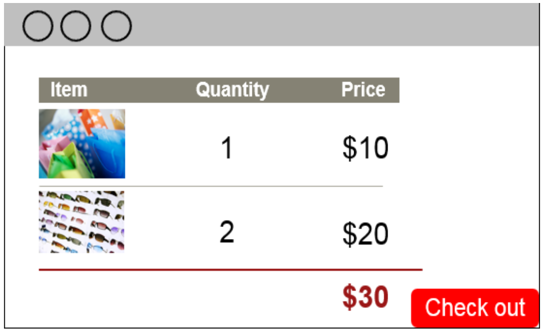
Earning while learning: Designing A/B tests

See Shahriari et al. (2015)

- Situation: test impact of different buttons on conversion rate.
- 2 variants $v \in \{1,2\}$
- v_i is variant shown to user i
- Binary outcome „buy“ versus „no buy“ $y_i \in \{0, 1\}$
- Data: $D_n = \{(v_i, y_i)\}_{i=1}^{n}$
- Objectives:
  - **Exploration**: explore both variants to learn which is more effective
  - **Exploitation**: earn higher profits using the more effective variant



vs.



Shiny…

Agenda

- About A.T. Kearney
- The idea behind Bayesian Optimization
- Earning while learning: A/B testing
- **Tuning algorithms with BO**
- Practical considerations

ATKearney

We helped a client to optimize their production and distribution footprint

Production and distribution footprint optimization

Food & Beverage

Client's situation and approach

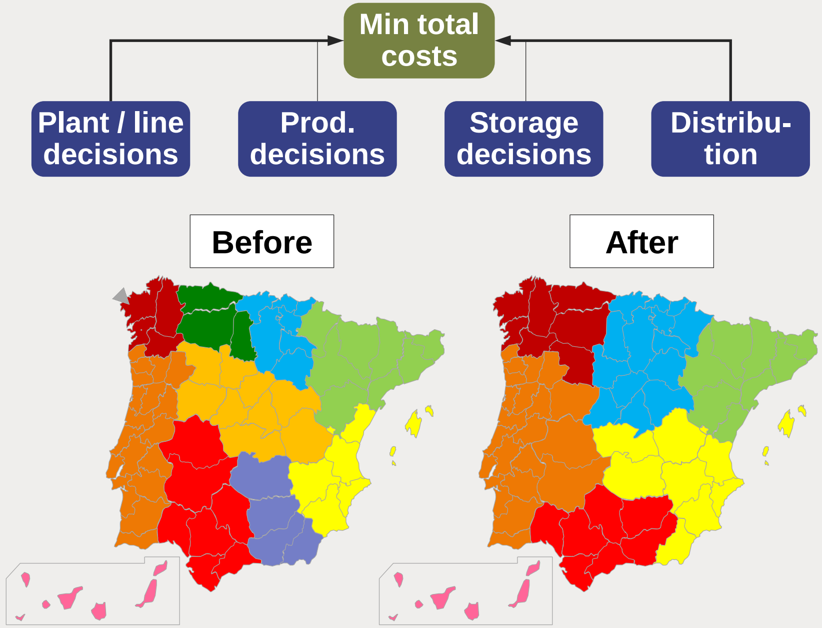
Situation

- The client's overall production network had significant overcapacity and production fragmentation
- The assignment of POS to warehouses was also not optimal and had to be optimized
- Our goal was to identify an optimal and robust footprint for the period 2015-2020

Approach

- We have implemented a **custom-built optimization model** of the production and distribution footprint
- The goal of the model to find most optimal overall network costs given the client's business constraints
- The implementation was done in AIMMS (www.aimms.com)
- AIMMS uses **IBM's ILOG CPLEX** under the hood

Optimization logic and outcomes



ATKearney

CPLEX has 76 parameters that need to be set

Algorithm	Parameter type	# parameters of this type	# values considered	Total # configurations
CPLEX MILP (MIQCP)	Boolean	6 (7)	2	$1.90 \cdot 10^{47}$ ($3.40 \cdot 10^{45}$)
	Categorical	45 (43)	3–7	
	Integer	18	5–7	
	Continuous	7	5–8	

[1]

"Integer programming problems are more sensitive to specific parameter settings, so **you may need to experiment** with them."

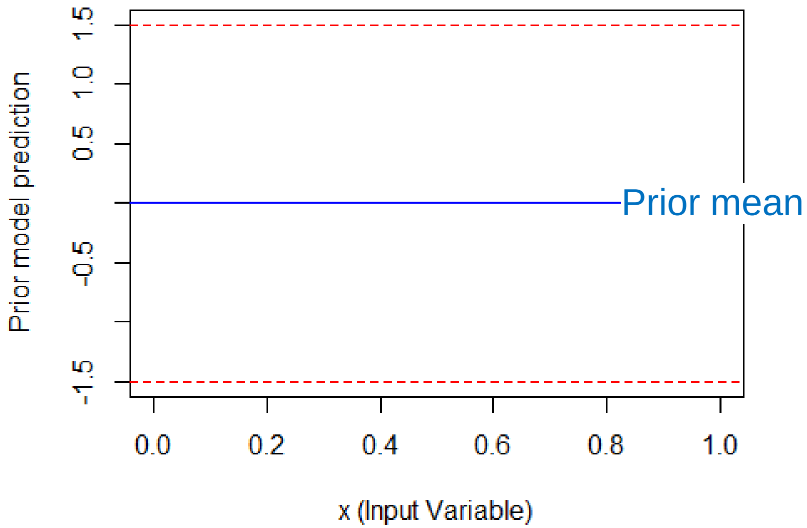
(ILOG CPLEX 12.1 user manual, p. 235)

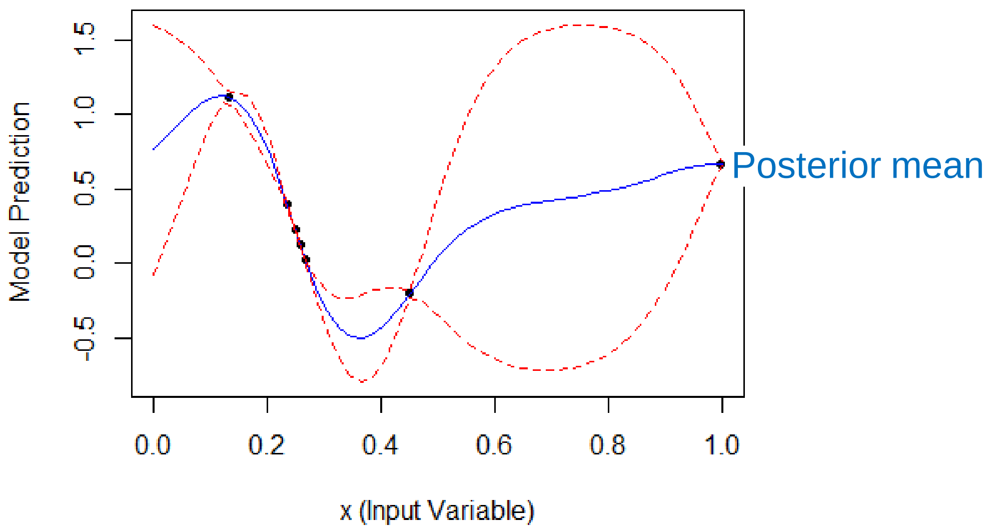
Number of possible configurations prohibitive to evaluate completely

1. Table taken from Hutter et al. (2010)
Source: A.T. Kearney

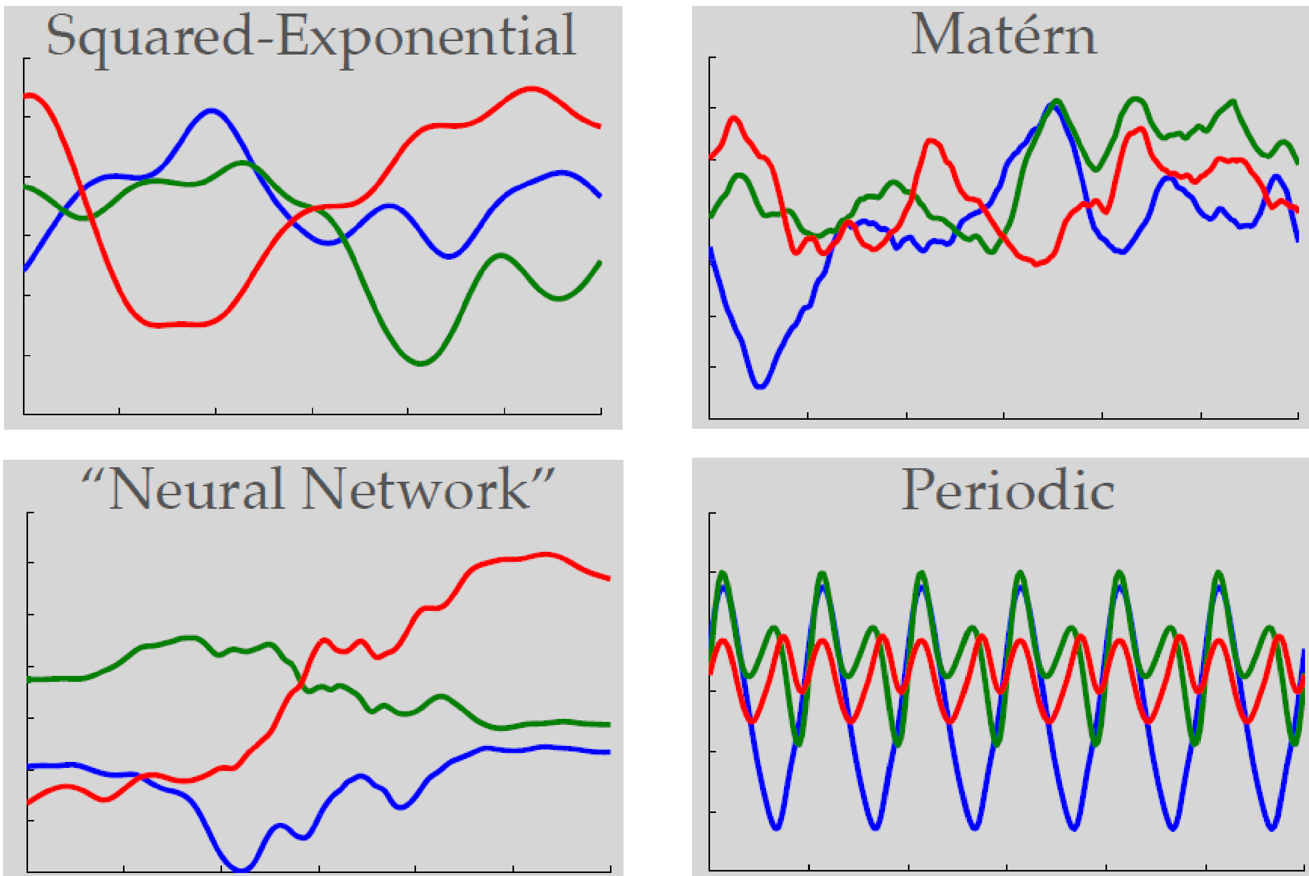
Bayesian optimization with Gaussian Processes

Gaussian Process as prior over functions





Common prior beliefs for the functional form



1. Source: Ryan Adams, A Tutorial on Bayesian Optimization for Machine Learning.

Shiny…

Agenda

- About A.T. Kearney
- The idea behind Bayesian Optimization
- Earning while learning: A/B testing
- Tuning algorithms with BO
- **Practical considerations**

And in practice…?

Works well over a small domain, but tricky for real problems.

- There is no standard software package to do this
  - Whetlab acquired by Twitter (now Spearmint [Python] et al.)
  - ParamILS and FocusedILS [both Java] to optimize IBM CPLEX

- Some important modelling choices in the background
  - Performance quite sensitive

And in practice…?

What are tricks to bring this to the big data domain

- Updating GP prior computationally expensive, problem in high dimensions
  - Idea: in practice objective function often ‚flat' along several dimensions, i.e. problem is actually lower-dimensional
  - Trick: dimension reduction through random embeddings (REMBO, Wang et al., 2013).

- Recursive algorithm: how to parallelize?
  - Idea: not only decide on the next point, but the next n points
  - Trick: use expected acquisition function, then distribute task

- Cut unpromising runs:
  - E.g. in ML: in case test error in inner loop is high, abort early to save computing resources.

ATKearney

A.T. Kearney is a leading global management consulting firm with offices in more than 40 countries. Since 1926, we have been trusted advisors to the world's foremost organizations. A.T. Kearney is a partner-owned firm, committed to helping clients achieve immediate impact and growing advantage on their most mission-critical issues. For more information, visit www.atkearney.com.

Americas	Atlanta Bogotá	Calgary Chicago	Dallas Detroit	Houston Mexico City	New York Palo Alto	San Francisco São Paulo	Toronto Washington, D.C.
Asia Pacific	Bangkok Beijing	Hong Kong Jakarta	Kuala Lumpur Melbourne	Mumbai New Delhi	Seoul Shanghai	Singapore Sydney	Taipei Tokyo
Europe	Amsterdam Berlin Brussels Bucharest	Budapest Copenhagen Düsseldorf Frankfurt	Helsinki Istanbul Kiev Lisbon	Ljubljana London Madrid Milan	Moscow Munich Oslo Paris	Prague Rome Stockholm Stuttgart	Vienna Warsaw Zurich
Middle East and Africa	Abu Dhabi Doha	Dubai Johannesburg	Manama Riyadh				