Gaussian Models for Fusion of Data from Different Sources

E. Burnaev
Skoltech

Multi-Fidelity Data Modeling: Motivation

Example 1

Adaptive Multi-Fidelity Active Search in Information Networks

Problem statement

- The data are represented in the form of Information Network: a graph that has associated attributes/labels with vertices or edges
- The user wants to find vertices relevant to the set of provided examples
- The system can ask the user to evaluate additional vertices or do relevance estimations via other sources and methods
- The goal is to achieve maximum payout within a limited budget

Challenges

- Each evaluation operation (both by user and system) has a cost (time/money)
- The budget is very limited
- The system has to decide the evaluation method (user/internal) and which vertices have to be evaluated

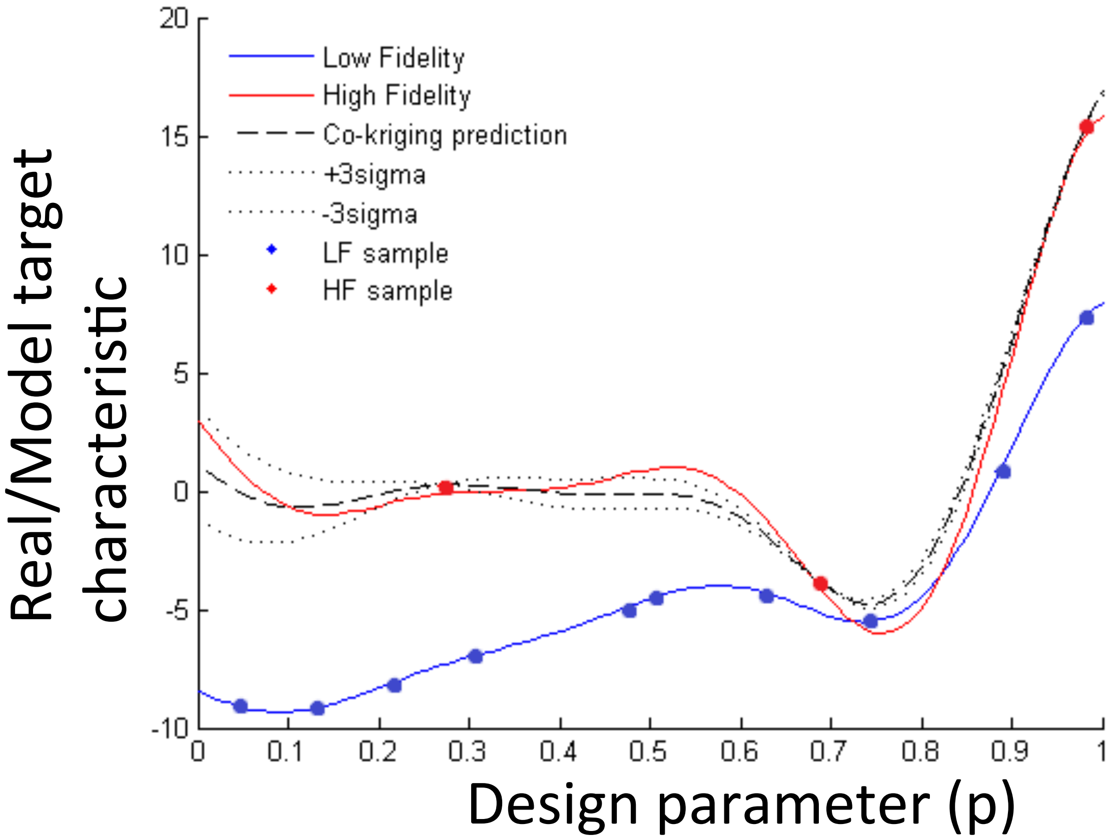
Multi-fidelity Surrogate Modeling

Low-fidelity model

High-fidelity model

$$Real(p) = \rho * Model(p) + \delta(p)$$

Example: parameter optimization with MFSM



Idea

Low-fidelity model

System's score

- Objectified
- Cheap
- Biased

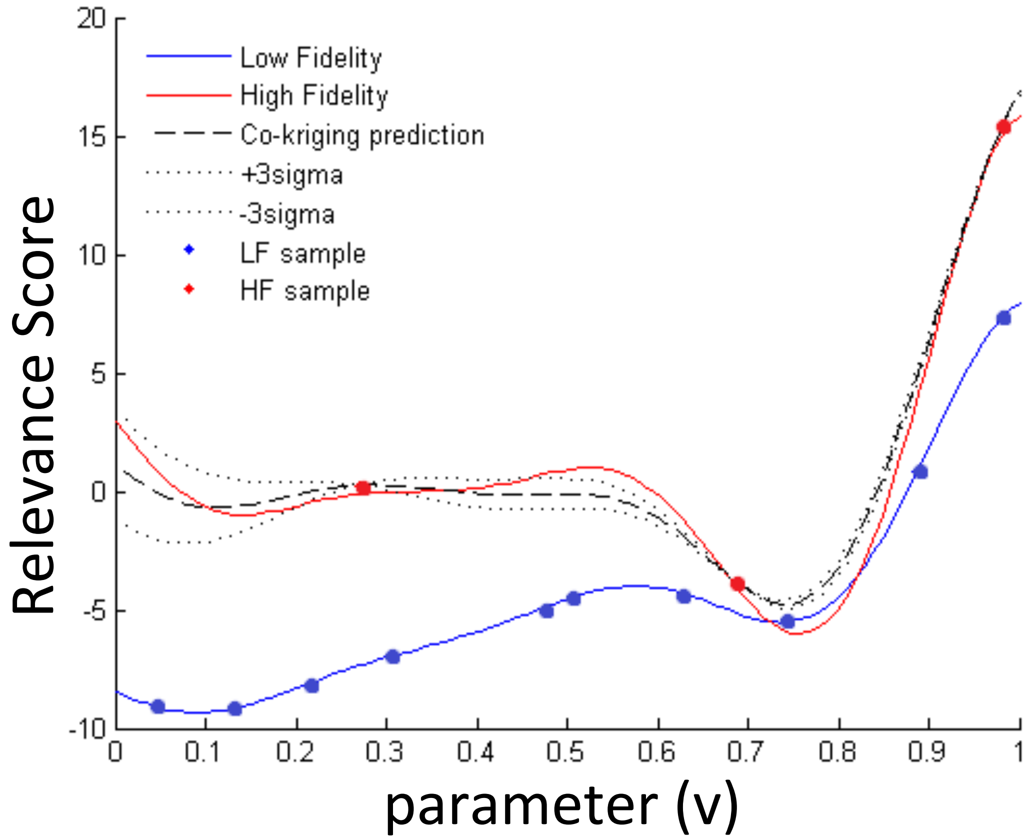
High-fidelity model

User's score

- Personalized
- Expensive
- More accurate

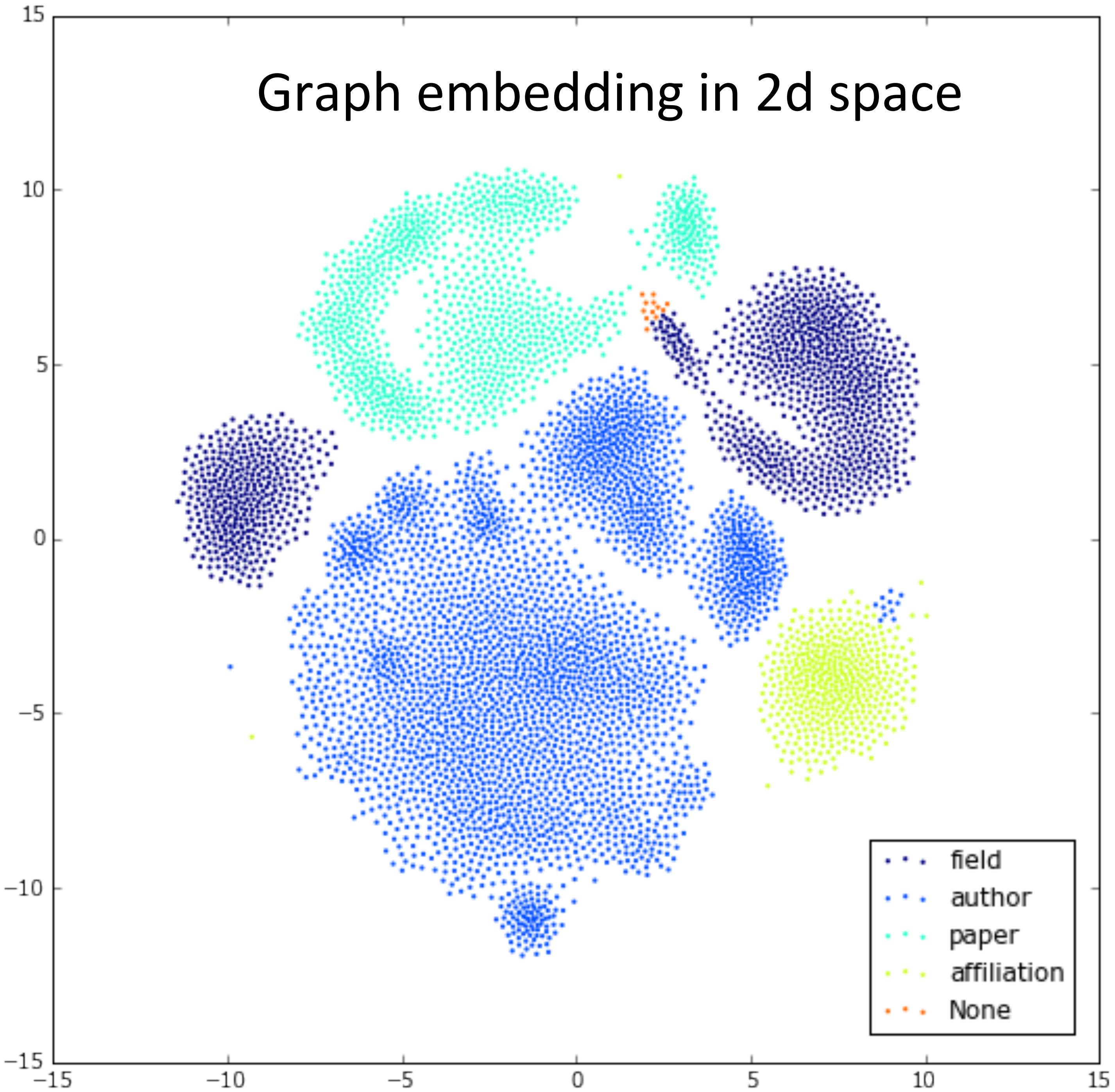
$$UserScore(v) = \rho * SytemScore(v) + \delta(v)$$

Example: parameter optimization with MFSM



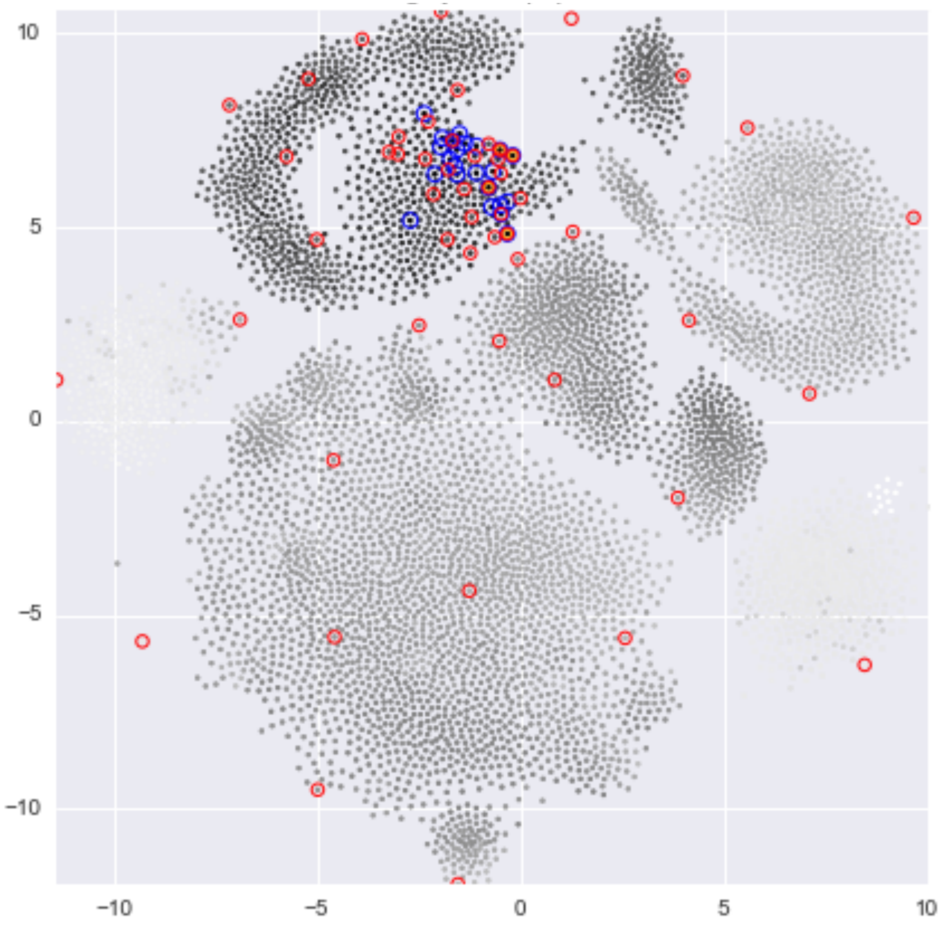
Example data:

- Microsoft academic graph
- 6K vertices (from KDD conference since 2012)
- Vertices types: papers, authors, fields, affiliations
- Edges types: 'hasAffiliation', 'hasField', 'isAuthor', 'isCitedBy', 'isSubfieldOf'

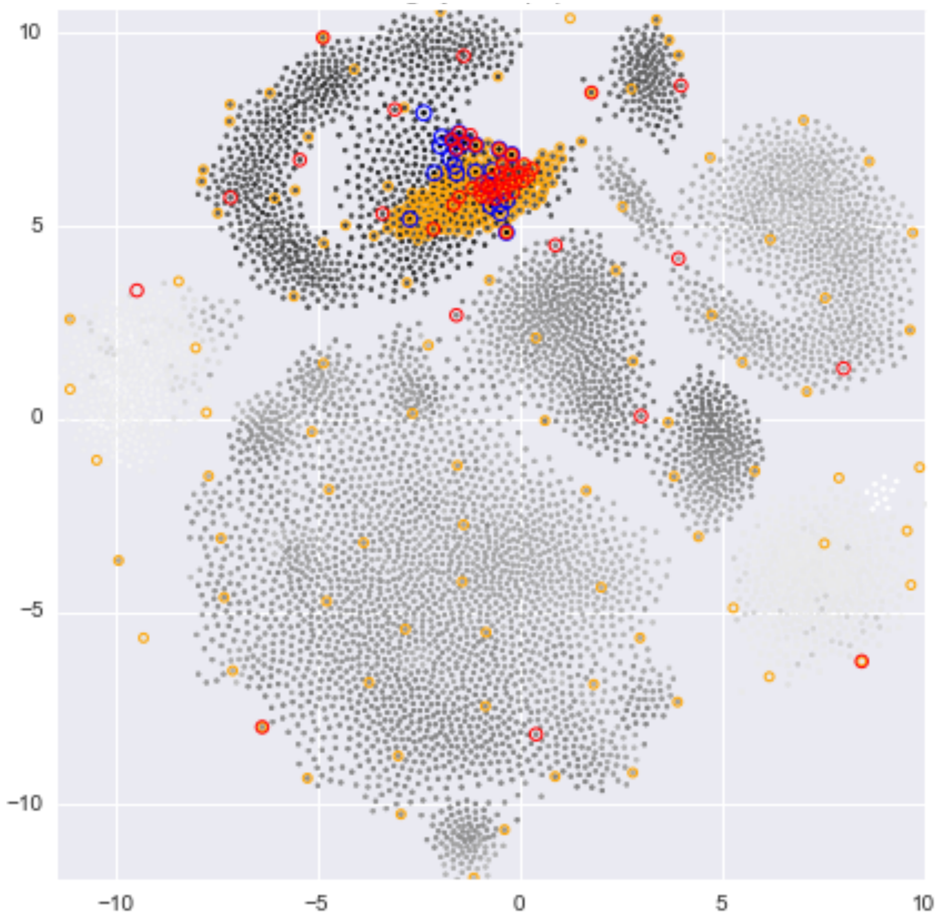


Example target 1: Highly cited papers

Using high-fidelity only

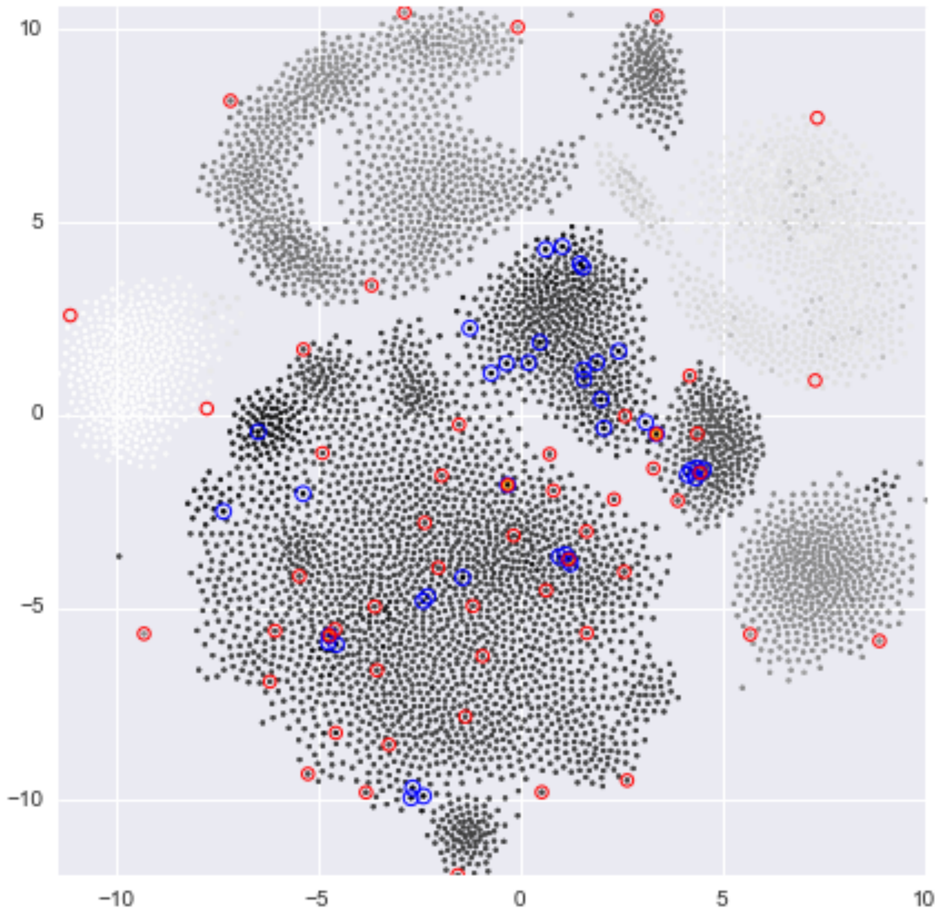


5 low-fidelity evaluations per 1 high-fidelity

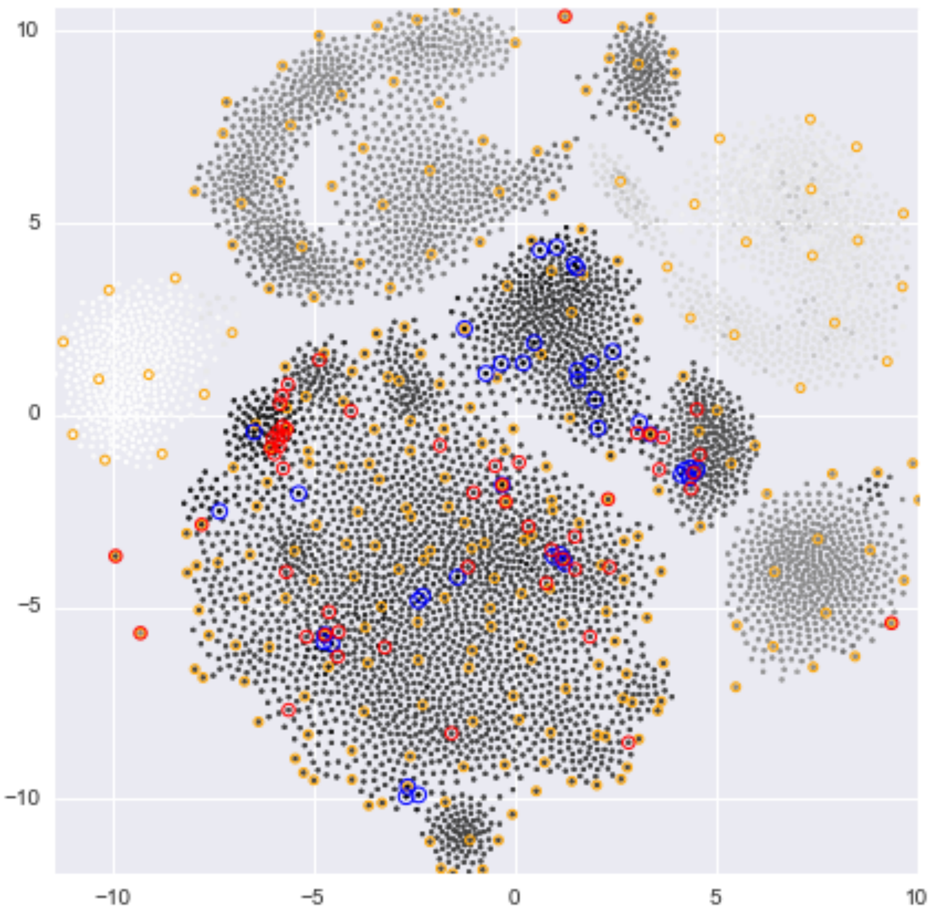


Example target 2: Authors in the genomics field

Using high-fidelity only

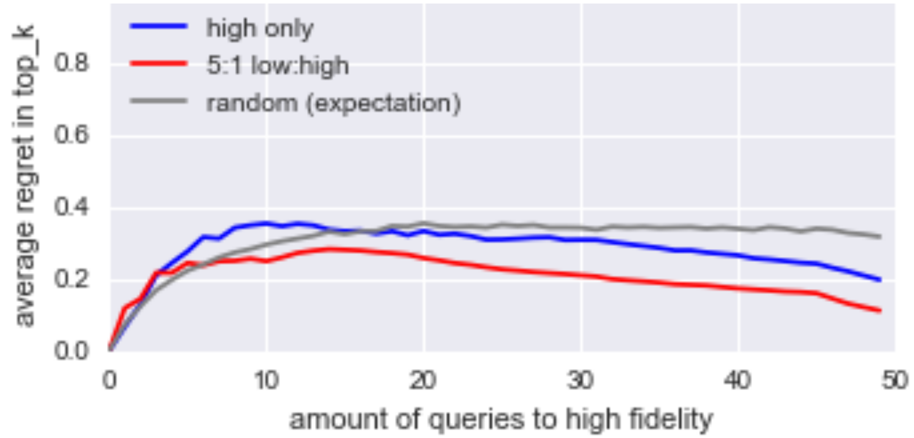


5 low-fidelity evaluations per 1 high-fidelity

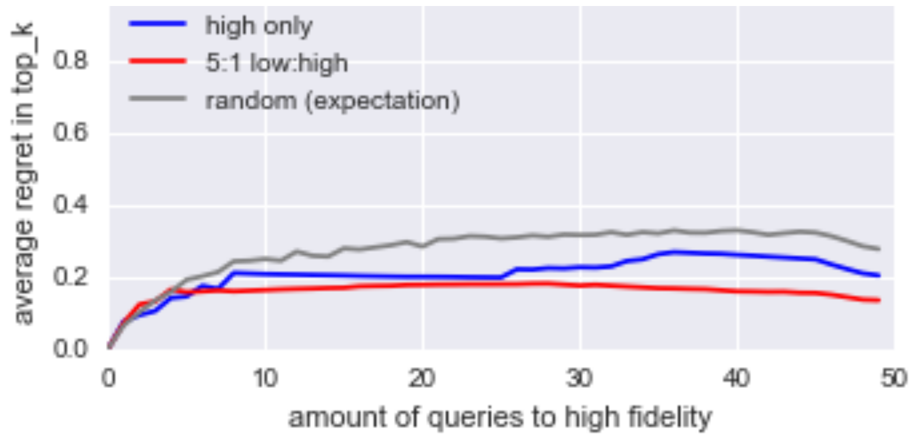


Regret measurements

Example 1

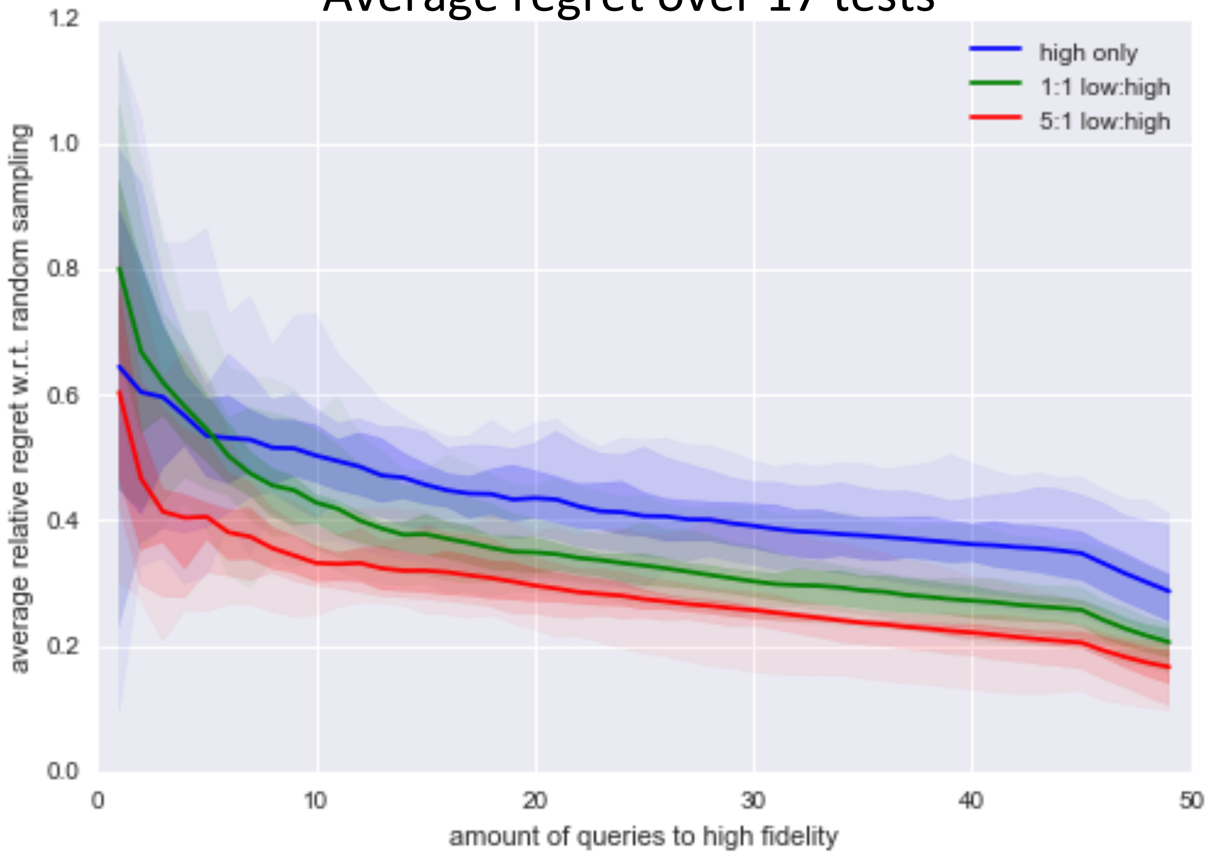


Example 2

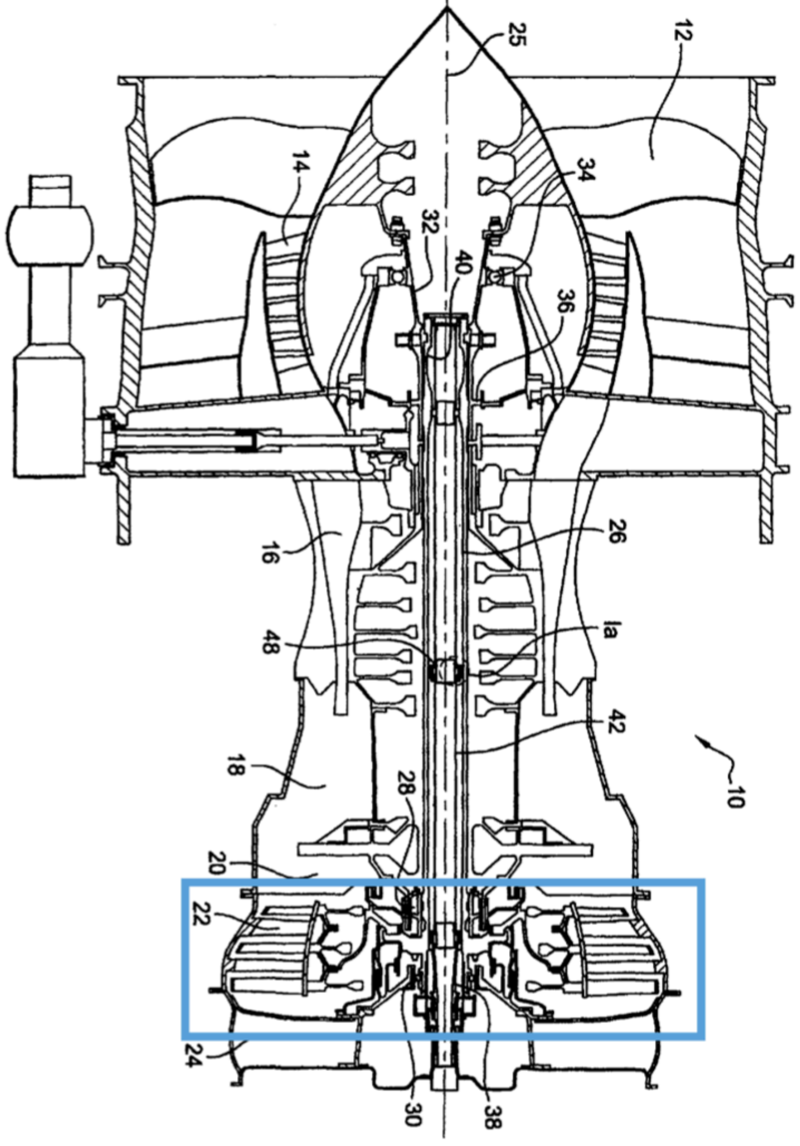


$$AvgReget_k = \frac{Best_k - Evaluated_k}{k}$$

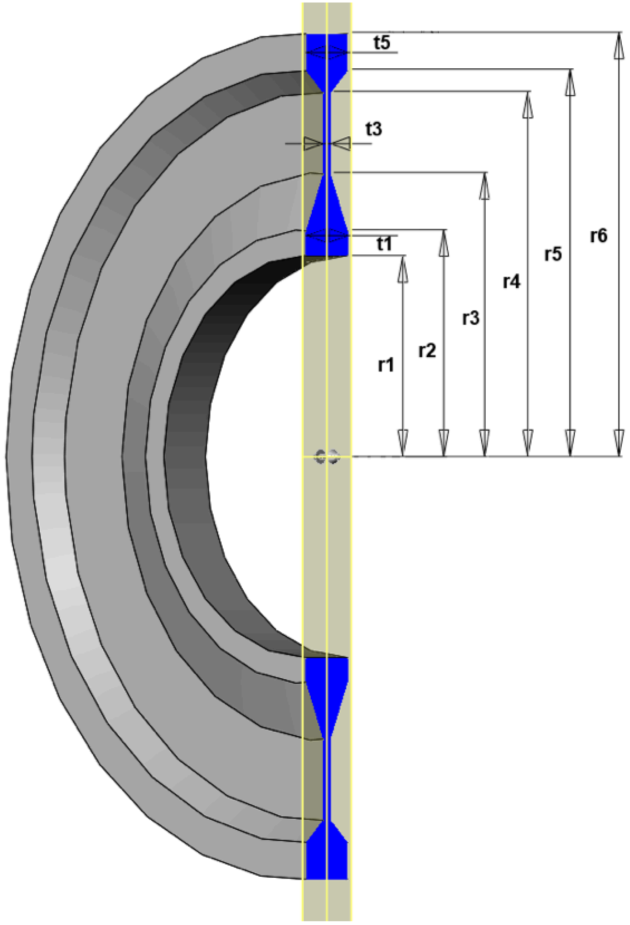
Average regret over 17 tests



Example 2
Engineering Optimization of Rotating Disk Shape



(a) Aircraft engine. Rotating disk is shown by the bold rectangle at the right side of the figure



(b) Rotating disk geometry

Fig. 1: Rotating disk problem

- Maximum radial displacement $u_{\max}$ and maximum stress $s_{\max}$ describes quality of a rotating disk.
- $y_l(\mathbf{x})$ is an Ordinary Differential Equations (ODE) solver.
- $y_h(\mathbf{x})$ is a Finite Elements Model (FEM) solver.
- We construct a model:
  - Parametrization of geometry of a rotating disk is input $\mathbf{x}$,
  - $u_{\max}$, $s_{\max}$ are two outputs $\mathbf{y}(\mathbf{x})$.

Model	Fidelity	CPU time (s)	Sample size
ODE	Low	0.01	5000
FEM	High	300	100

$$
\begin{aligned}
m, u_{\max} \to \min_{r_1,\ldots,r_6,t_1,t_3,t_5}&, \\
u_{\max} \le 0.3, s_{max} \le 600&, \\
10 \le r_1 \le 110, 120 \le r_2 \le 140&, \\
150 \le r_3 \le 168, 170 \le r_4 \le 200&, \\
4 \le t_1 \le 50, 4 \le t_3 \le 50&, \\
r_5 = 210, r_6 = 230, t_5 = 32&.
\end{aligned}
$$