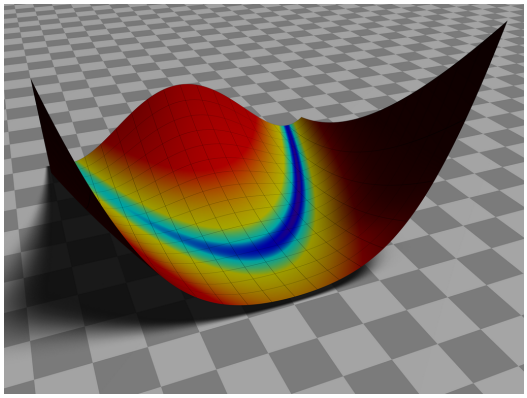
Sampling

Lecture 31



ME EN 575
Andrew Ning
aning@byu.edu

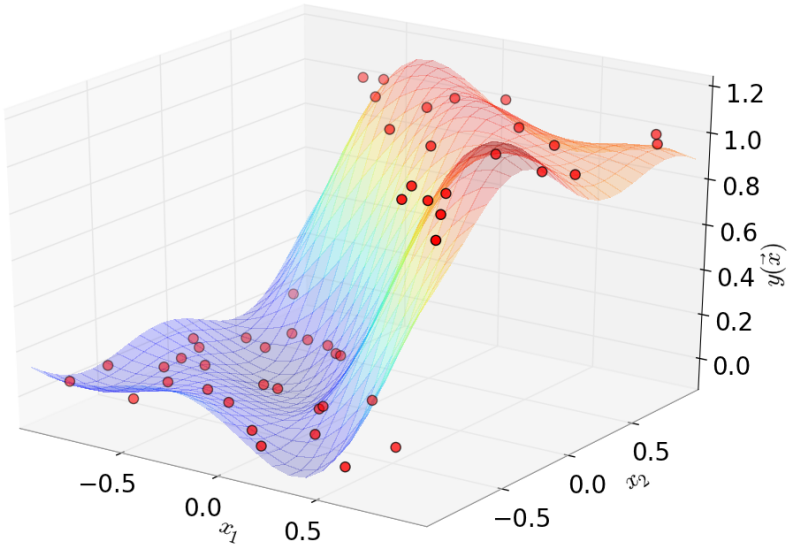
Outline

Surrogate Based Optimization (SBO) Introduction

Sampling

Surrogate Based Optimization (SBO) Introduction

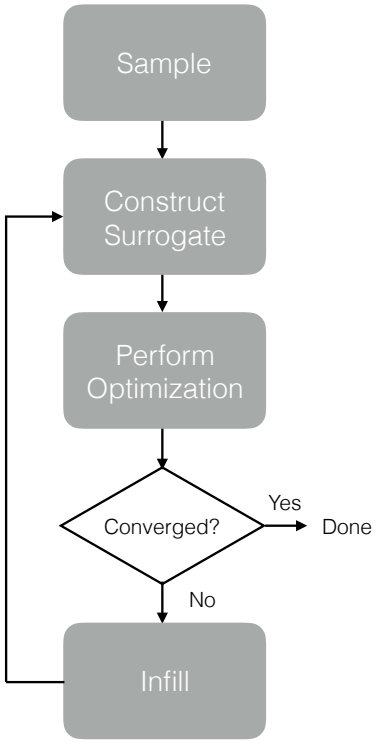
What is a surrogate model?



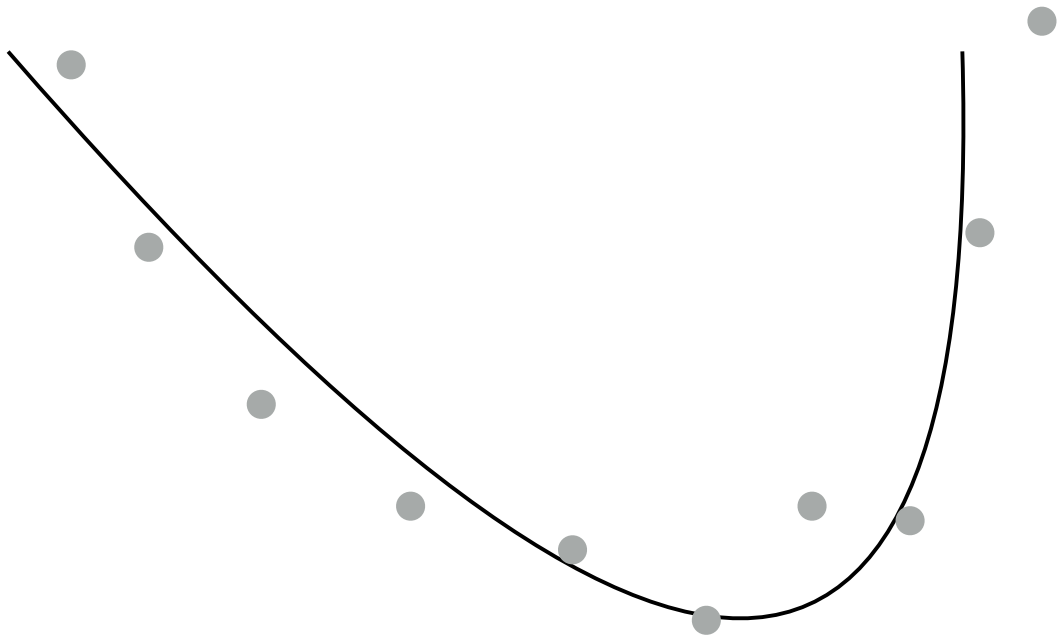
When might one use a surrogate model?

- Simulation is expensive (takes a long time to run).
- Simulation output is noisy.
- Experimental data.
- Understand functional relationships
- Multifidelity optimization.

Procedure



Sampling



What if you have 10 variables?

Even with 10 points in each dimension that's 10^10 posible combinations.

If each simulation takes one hour it would take over a million years to evaluate all of them.

We need to identify the most important variables.

There are many advanced methods we won't have time to get into.

One simple method: evaluate the gradient at many points then compute mean and standard deviation. If mean or standard deviation of g_i is large, then variable i may be important (large mean implies large effect on objective, large variance implies lots of interactions or highly nonlinear).

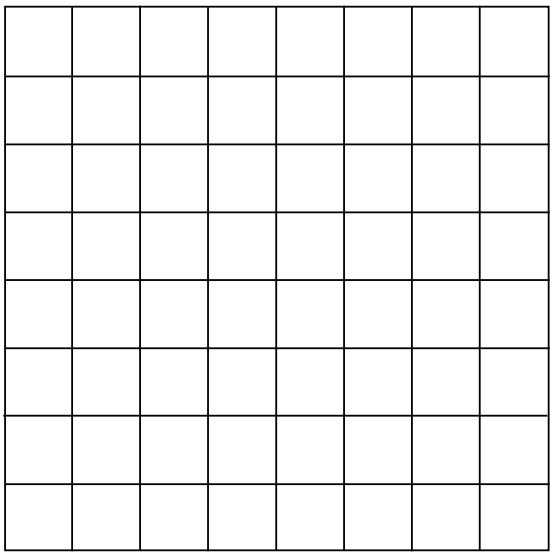
Even with a reduced variable set, we still need to choose where to sample carefully.

Full grid searches don't scale.

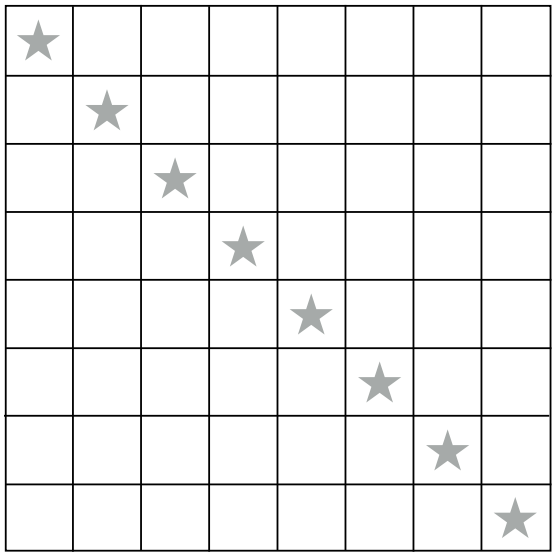
Random searches don't scale.

Latin Hypercube Sampling

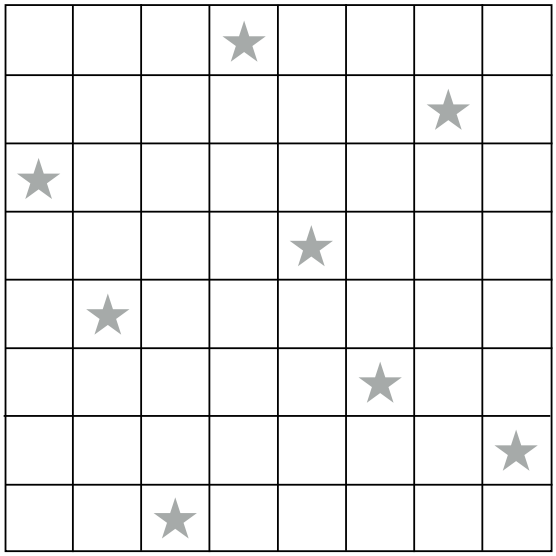
Let's say we have 2 variables, and want to take 8 samples. We could think about discretizing our space into 8 bins in each direction.



We should try to place one sample in each row and each column (this is called a Latin square and the higher dimension extension a Latin hypercube).



We also need our points to be space filling:



This is an optimization problem:

maximize: spread subject to: projection of samples on each axes follows specified probability distribution (uniform shown above but can work with any).

There are many solutions so there is some randomness. The procedure is called Latin Hypercube Sampling.

Where else would LHS be useful?

- Matlab: `lhsdesign` and `lhsnorm` (Statistics Toolbox)
- Python: `lhs` (pyDOE). Using the ppf function in scipy.stats you could use this with any distribution.

Matlab: lhsdesign demo.

Python: see Monte Carlo notebook from before, but this time with LHS.