Experimental optimization

Lecture 9: Response surface methodology

David Sweet

Review: Experimental cost

- Experiments are expensive:
  - They take time to run
  - They put users at risk of bad experience
  - They cost money: engineer’s salary, lost revenue
- The only way to reduce this cost is to **take fewer measurements**.
- Research into experimental methods seeks to reduce the number of measurements required to achieve an experiment’s goal

Experimental optimization

- One view of experimental optimization is “optimization with an expensive objective”
- Optimization generally: find the highest-value parameter or configuration
- Experimental optimization: optimize when “value” (business metric) is determined by an experiment; experiments are expensive

Specialization

Reduce cost for special cases

- One way to reduce costs is to create methods that take advantage of features of some subclass of systems
- Ex: Specialize to systems with a continuously-valued parameter
- Won’t work to optimize, ex., flags/booleans, categorical parameters discrete parameters; ex:
  - flags: use old ML model or new ML model
  - categorical: route orders to NYSE, NASDAQ, or BATS (exchanges)
  - discrete: Show the top K posts, where K in {1,2,3,4,5}

Specialization

Continuously-valued parameters

- Examples of continuously-valued parameters:
  - Trading threshold like “If E[return] > threshold, BUY”. Which threshold value maximizes profit?
  - Ranking signal weight: Order posts by signal_1 + w*signal_2. Which weight value maximizes number of posts viewed?
  - Fraud-detector threshold like “If P{fraud} > threshold, reject transaction”. Which threshold maximizes revenue?
- Any real value in some range would be an acceptable threshold or weight.

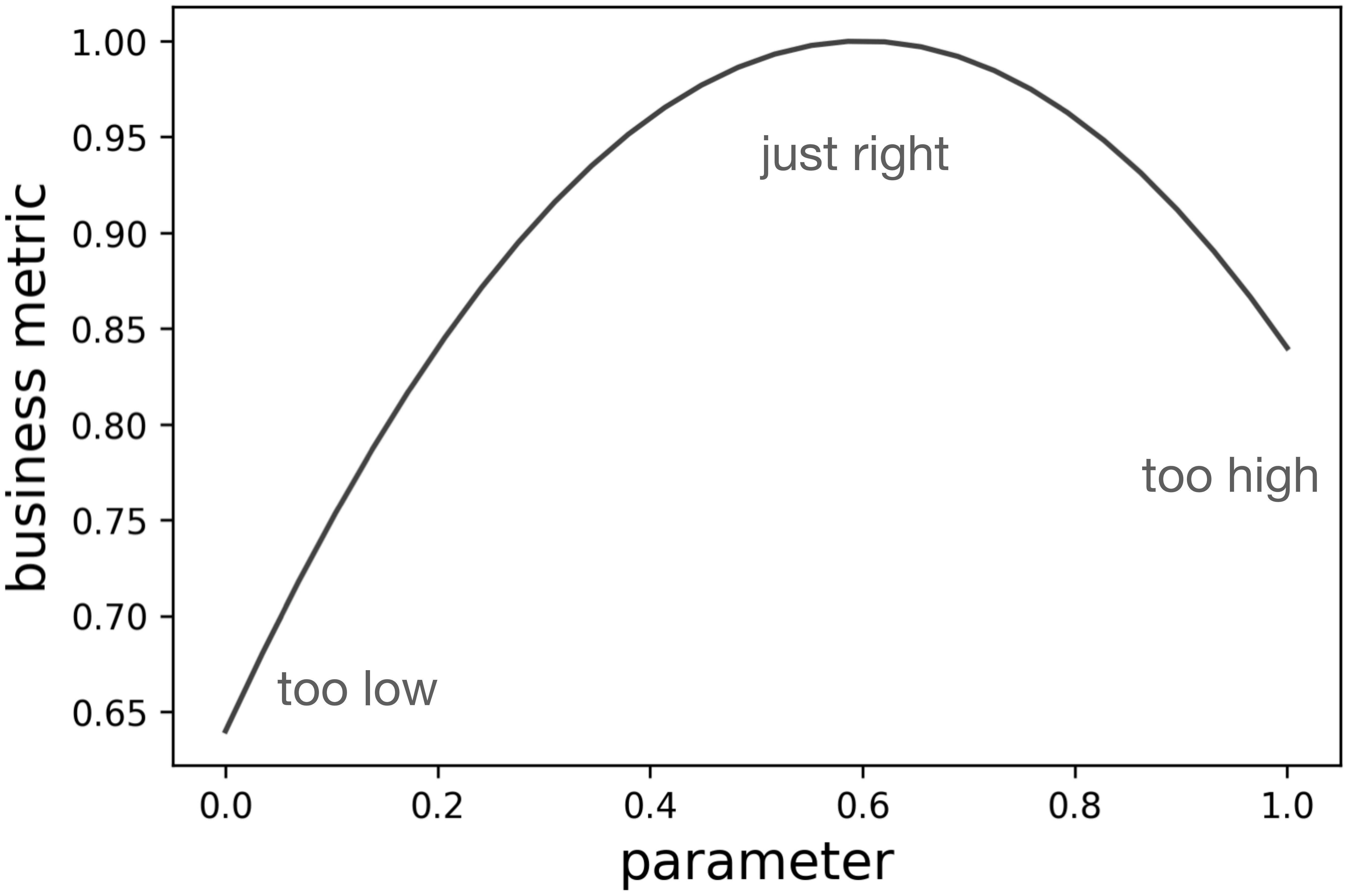
Optimizing continuous parameters
Competing concerns of business metric

- Trading threshold, profit = [# trades] * [profit/trade]
  - Too high, then too few trades.
  - Too low, then too little profit/trade.
- Ranking signal weight
  - Too high, maybe users spend time commenting & don’t scroll
  - Too low, maybe users just don’t like the posts they see
- Fraud-detector threshold:
  - Too high, too much revenue lost to fraud
  - Too low, too much revenue lost by rejecting good transactions

Optimizing continuous parameters

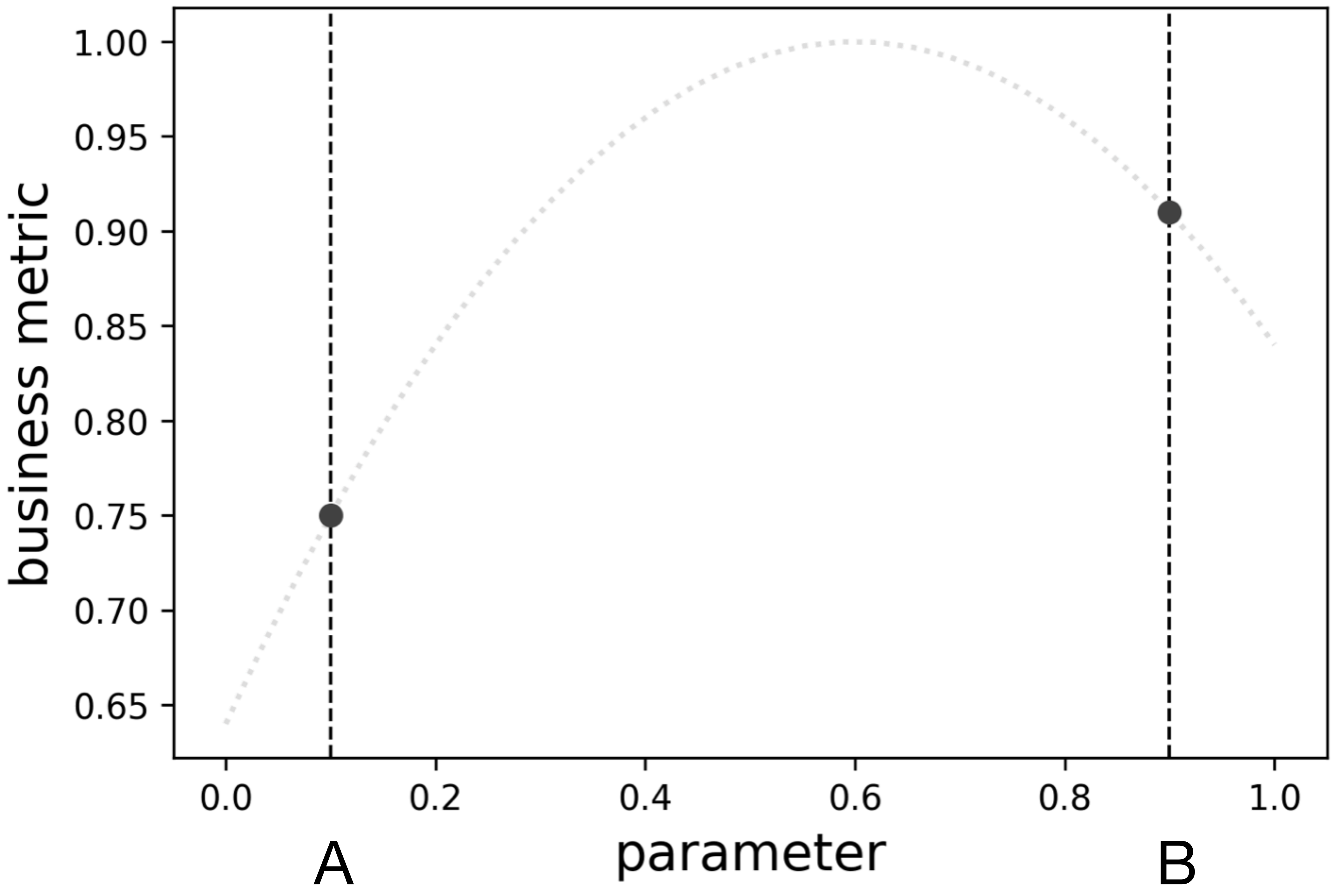
The response surface

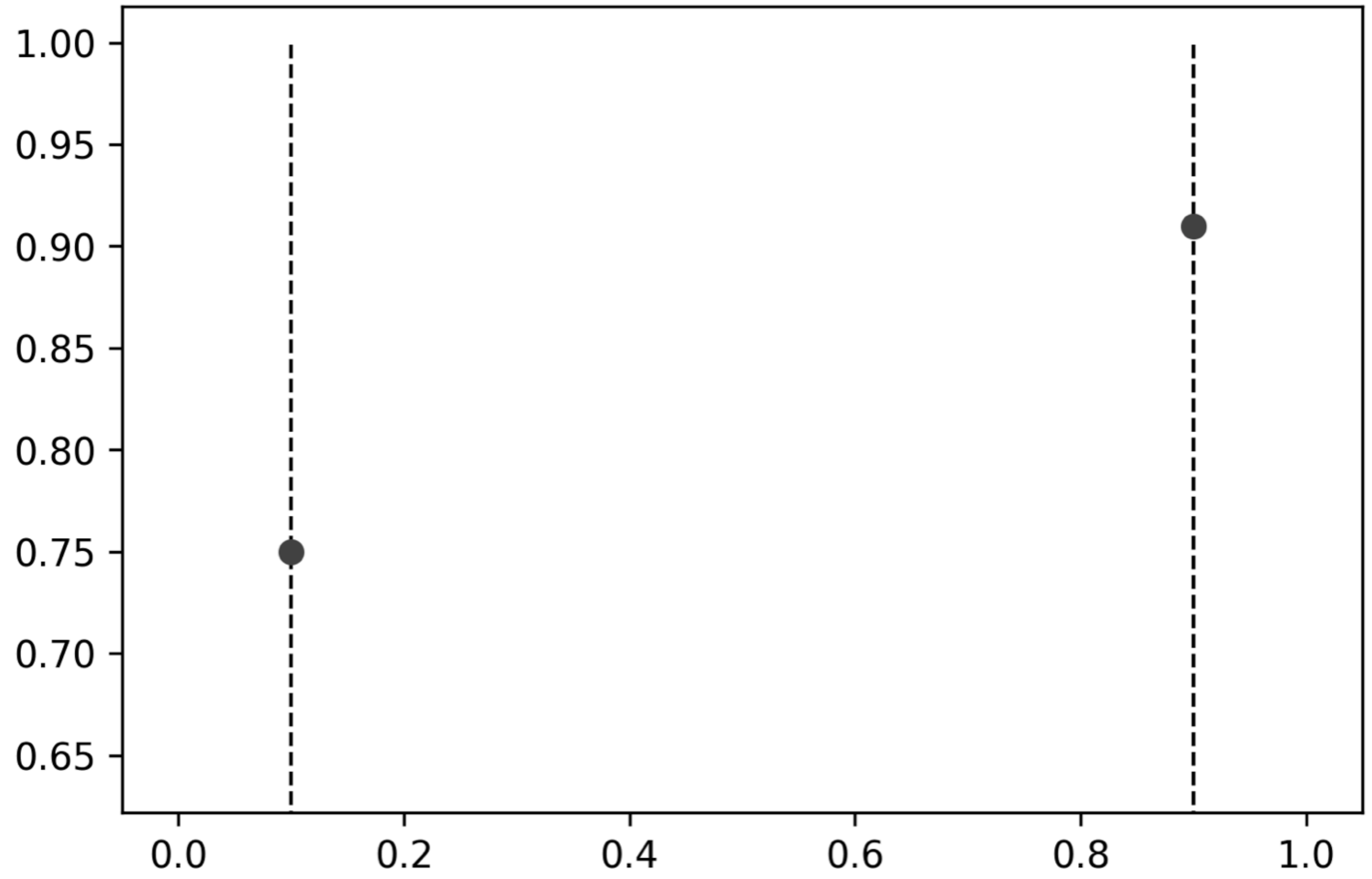
- Curve is called the *response surface* (RS)
- RS is the function "BM vs. parameter"
- Typically a "hump" because of competing concerns
- You can't observe the response surface.
- How could you find the maximum using experiments?



A/B testing
On a response surface

- Measure one low parameter value (A) and one high parameter value (B)
- Pick the better of the two
- This case: Choose B, higher BM
- B is not at the maximum.
- You could run more A/B tests, but how can you do that efficiently?
- You don't observe the response surface, only the dots

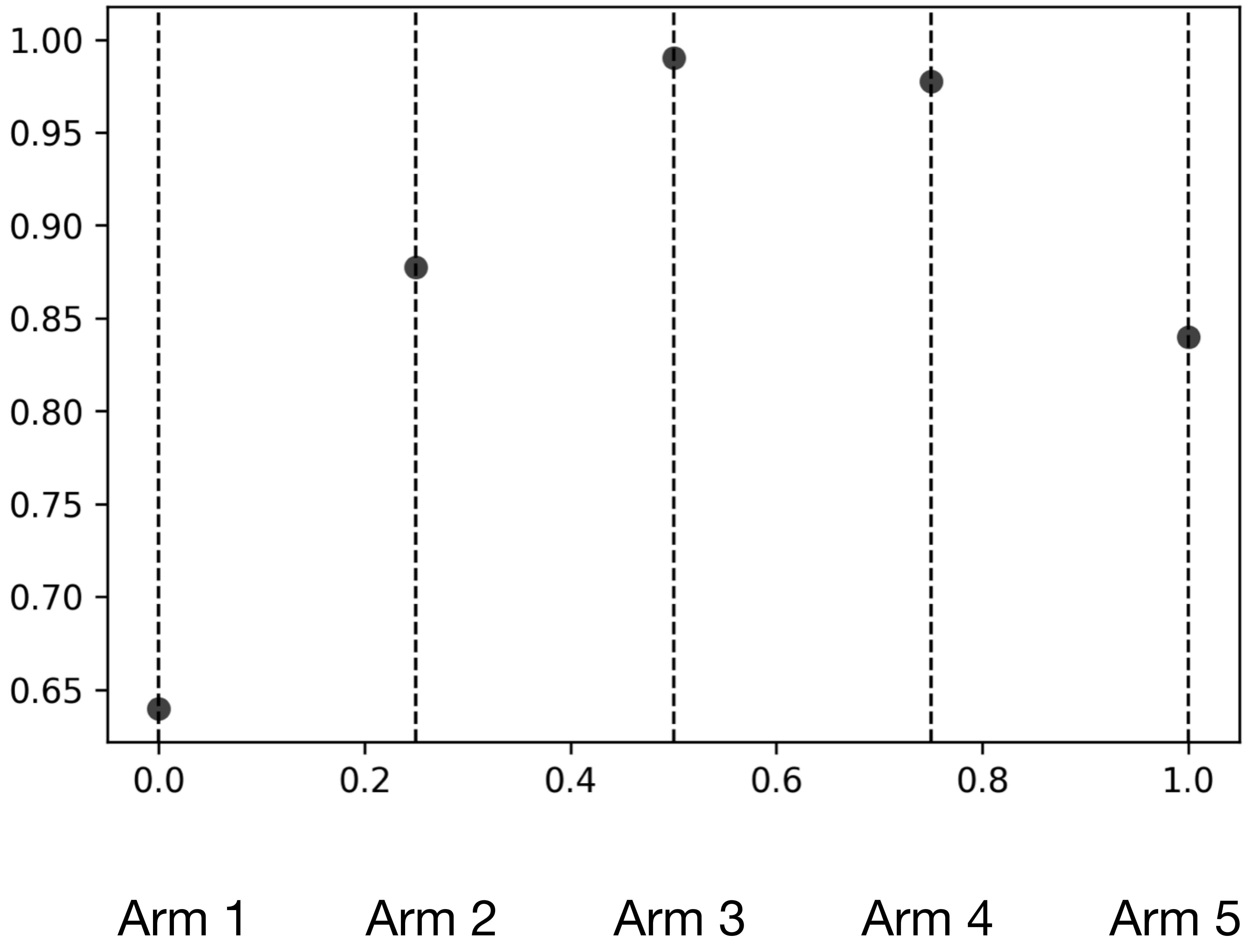




Multi-armed bandit

On a response surface

- Choose an evenly-spaced grid of parameter values as arms
- Run MAB to find best arm
- Want to get closer to maximum?
  - More arms
  - Longer experiment

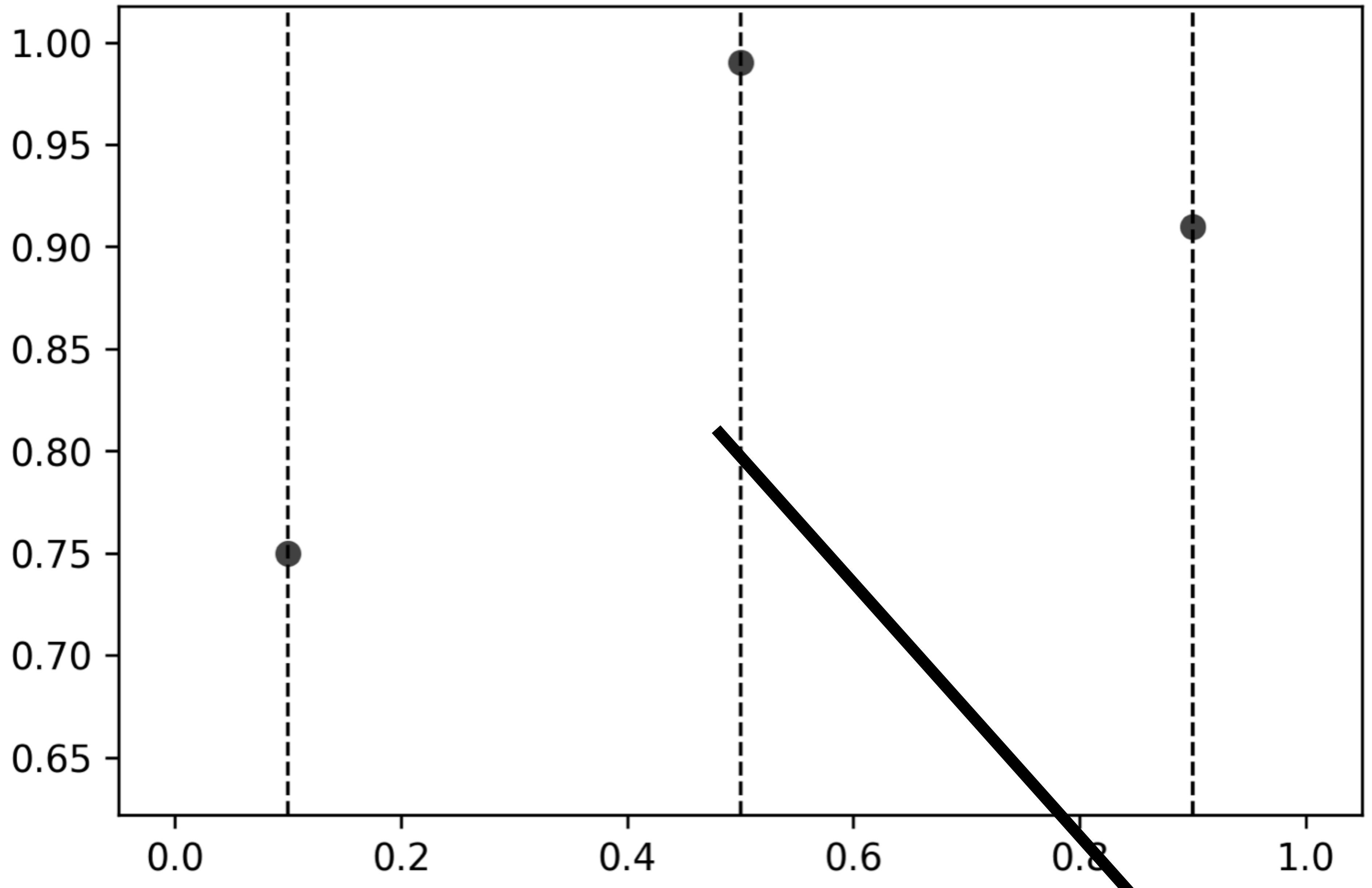


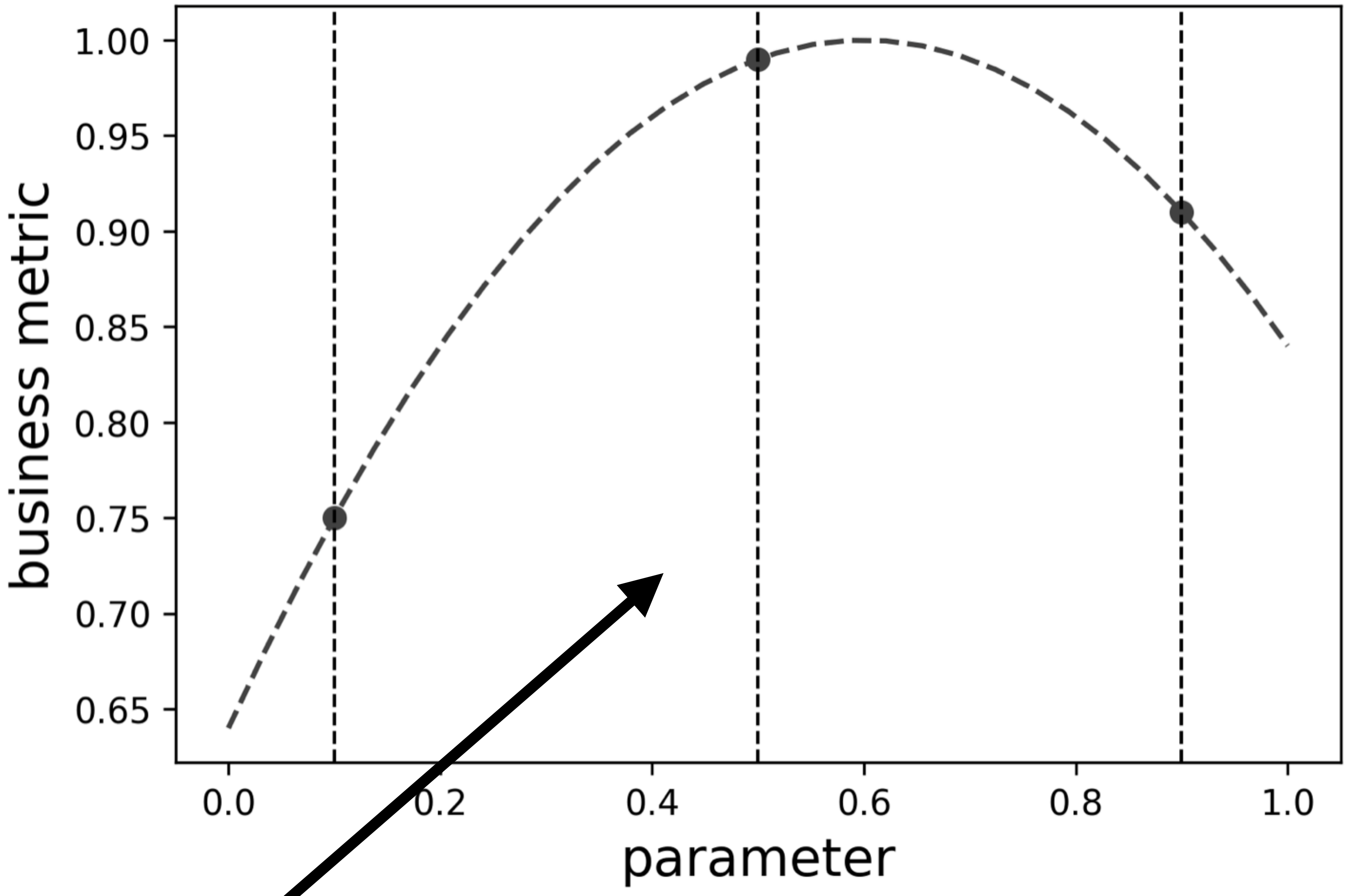
Response-surface methodology

Model the response surface directly

- Assert that the response surface (RS) has a hump
- Model the hump as a parabola
- Model is called a *surrogate function*
- You need three points to define a parabola
  - Two points define a line
- Take measurements at three parameter values

Response-surface methodology





Model as parabola

Response-surface methodology

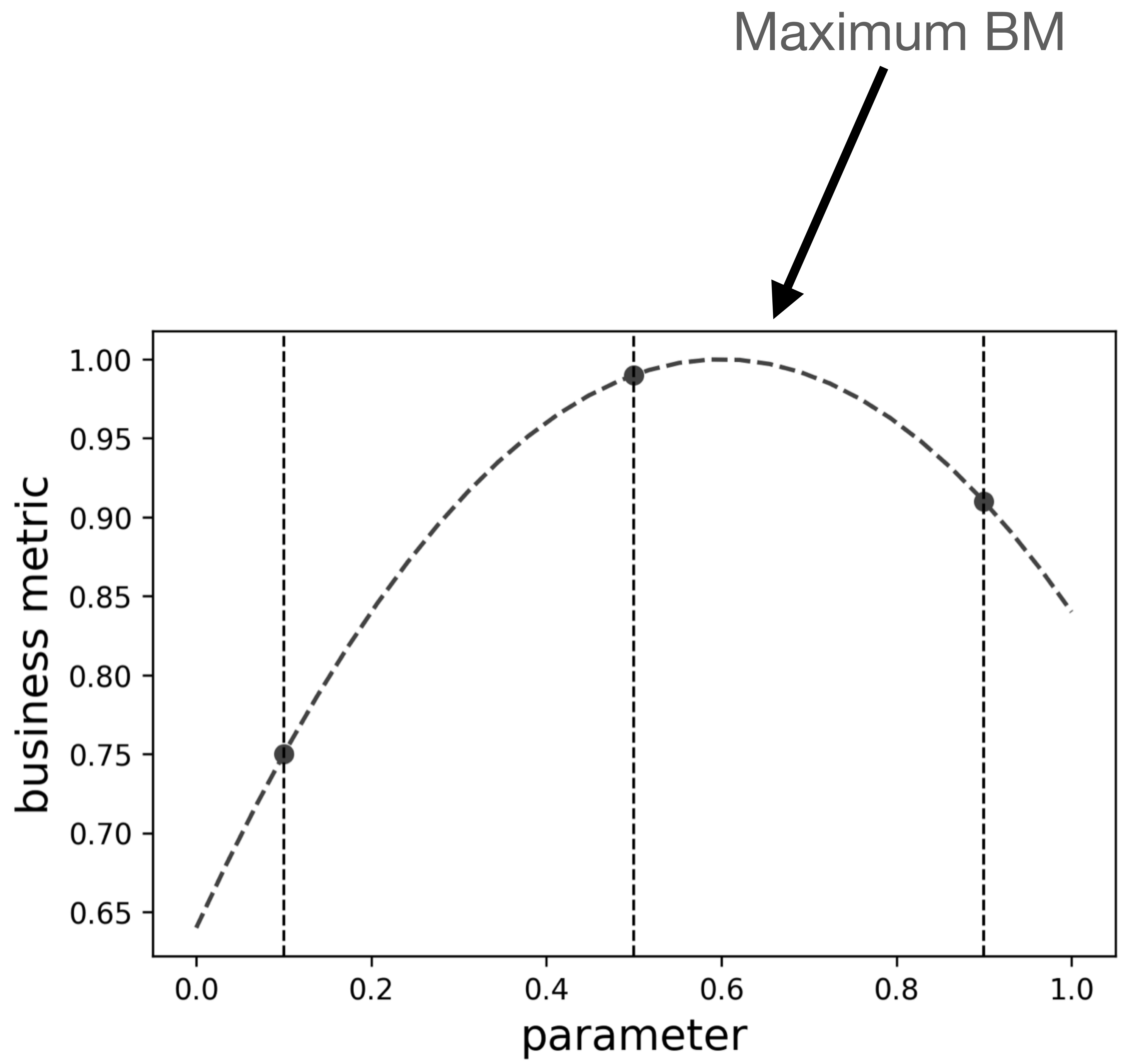
Model the response surface directly

- Three measurements of BM at low, medium, high parameter values
- Model as parabola: $BM = BM_0 - (parameter - parameter_0)^2$
- Or, with $y = BM, x = parameter$
- $y = y_0 + (x - x_0)^2$
- Equivalently: $y = ax^2 + bx + c$
- Use linear regression to find a, b, c

Response-surface methodology

Optimize over the model RS

- Now you can “see” the response surface, b/c you can plot your model parabola
- The maximum is at a value that you **did not measure**
- To find this maximum with A/B testing or MAB would have required many more measurements



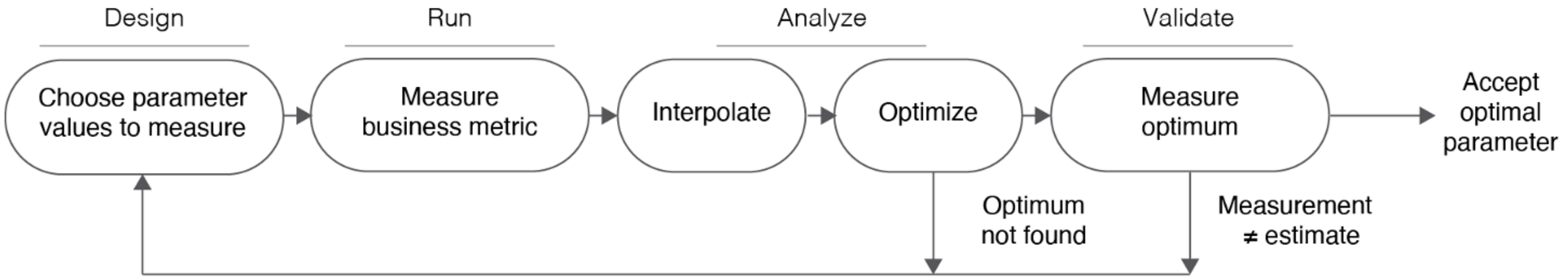
Response-surface methodology

Only an approximation

- Surrogate (model of RS) is only an approximation
- Validate by measuring at the predicted-best parameter
- You have collected four measurements.
- If you want an even better approximation, rebuild the model with all four measurements
    - Maybe shift the parameter range if the optimum is near the edge
- Then reoptimize, validate the new optimum, …
- Repeat until your predicted-optimum stops moving.

Response-surface methodology (RSM)

Summary of method



RSM measurements

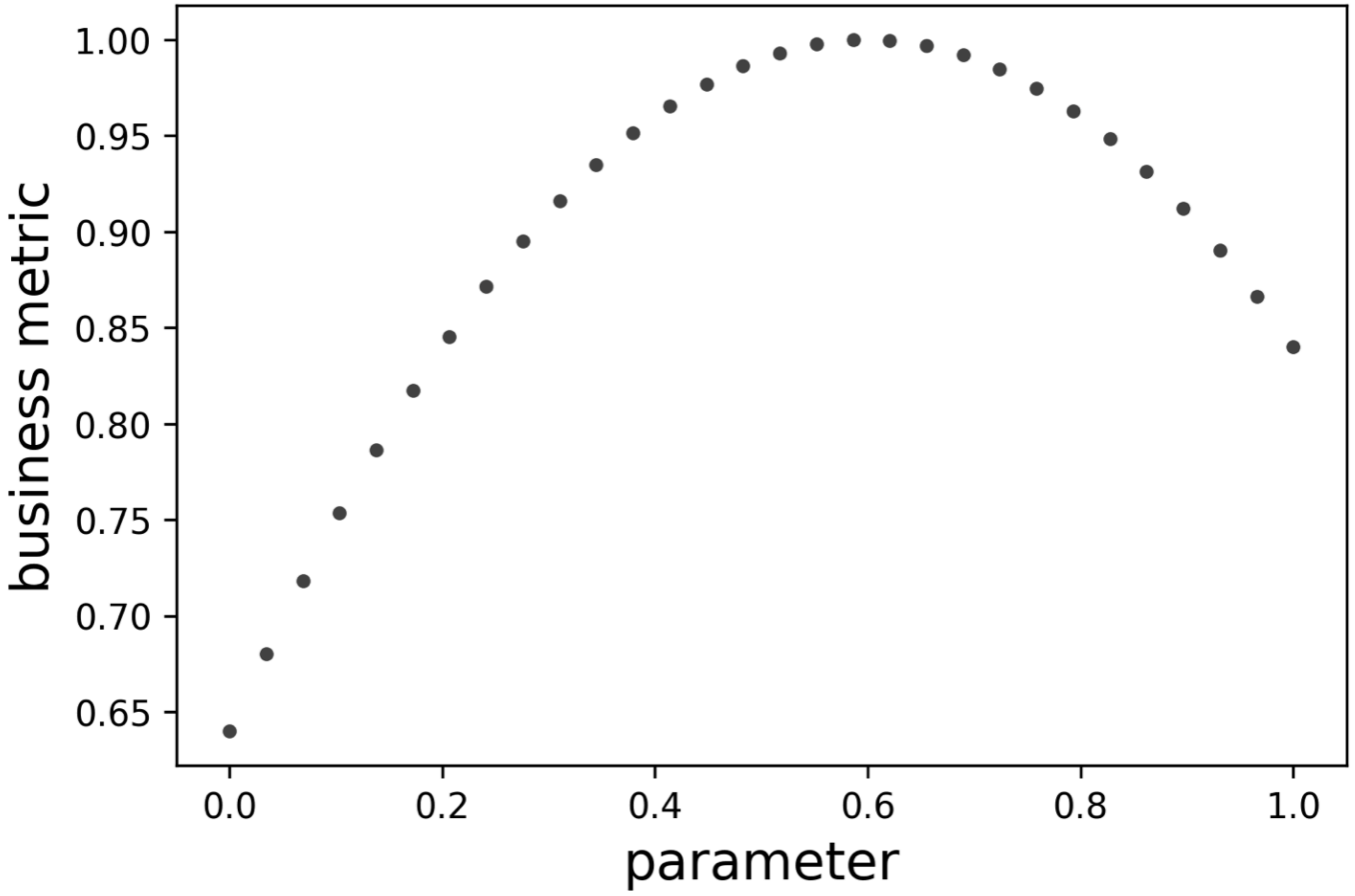
Use N from A/B testing

- RSM measurements are aggregate measurements
- Use $N = (\frac{2.48\hat{\sigma}}{PS})^2$
- PS here says “If the BM of two parameters is within PS, I’ll treat them as equivalent”
- Alternatively, “I want to be within PS of the true optimum”

RSM optimization

Grid search

- Optimization over the model is cheap, so evaluate many parameter values
- Use a grid search
- Pick the parameter with the highest model BM
- Ex: `np.linspace()`, `np.where()`



Response-surface methodology

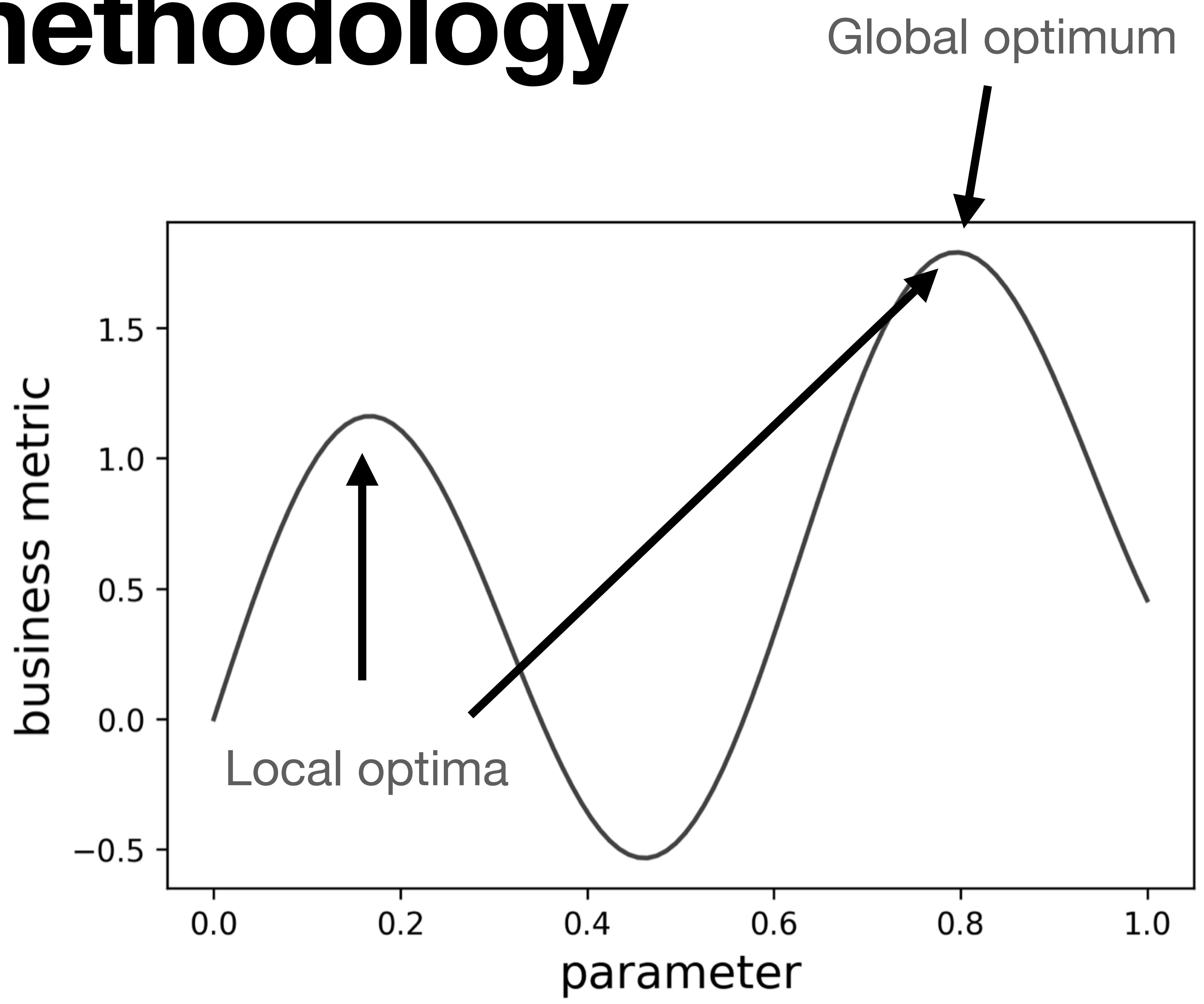
Multiple parameters

- Might want to optimize two thresholds or a threshold and a weight
- Could optimize separately but generally simultaneous optimization finds a better answer (see book appendix B)
- RSM works on 2 parameters — maybe even up to 5, with more advanced techniques
- We'll skip all that and learn Bayesian optimization instead
- Use RSM for one- or two-parameter problems where you'd prefer to visualize results and be "hands on" with the system
  - Ex., where you are uncertain about parameter ranges, concerned about system safety, unsure of tooling quality, or just in early stages of dealing with a system

Response-surface methodology

Local vs. global optima

- RS might have multiple humps
- You want the highest hump
- RSM will only search locally
- Think hard about parameter range
- Local optimum is better than no optimization at all!



Response-surface methodology

Summary

- Motivation: Reduce experimentation cost for continuously-valued parameters
- Surrogate function: Model the RS (BM vs. parameters) with linear regression
- RSM is interactive: You make decisions about where to measure and view visualizations
- RSM is iterative: You might repeat the design-measure-analyze loop multiple times