



האוניברסיטה  
העברית  
בירושלים  
THE HEBREW  
UNIVERSITY  
OF JERUSALEM

# Artificial Intelligence in Medicine

## Learning (1)

Nir Friedman and Tommy Kaplan

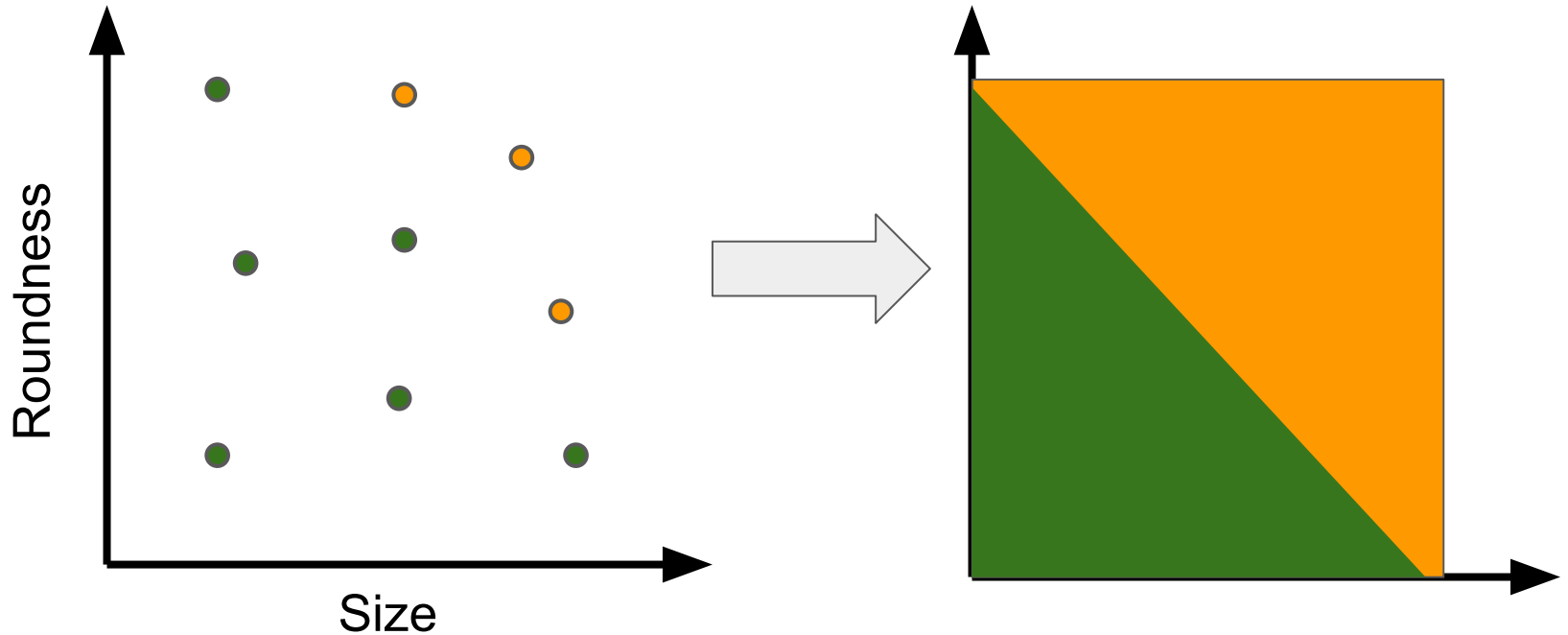
14/1/24

“It is a capital mistake to theorize  
before one has data.”

Sherlock Holmes



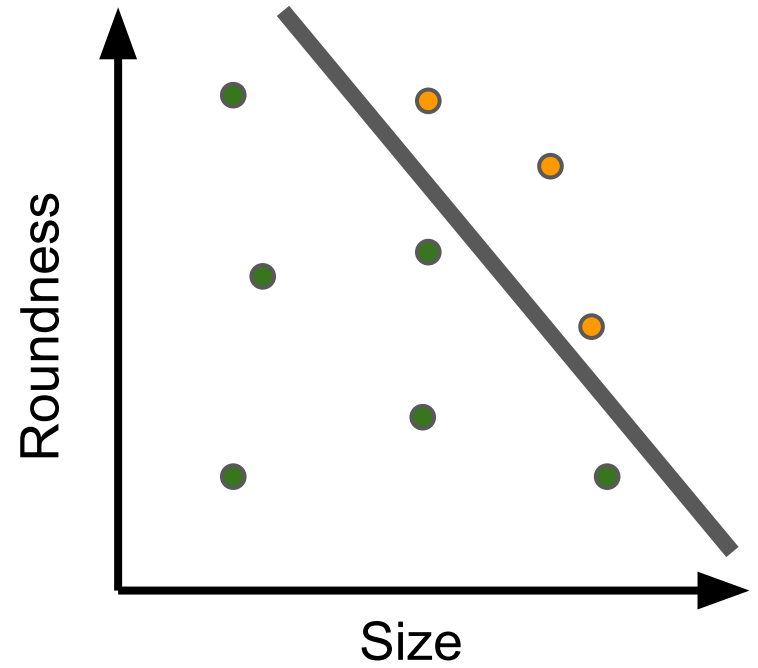
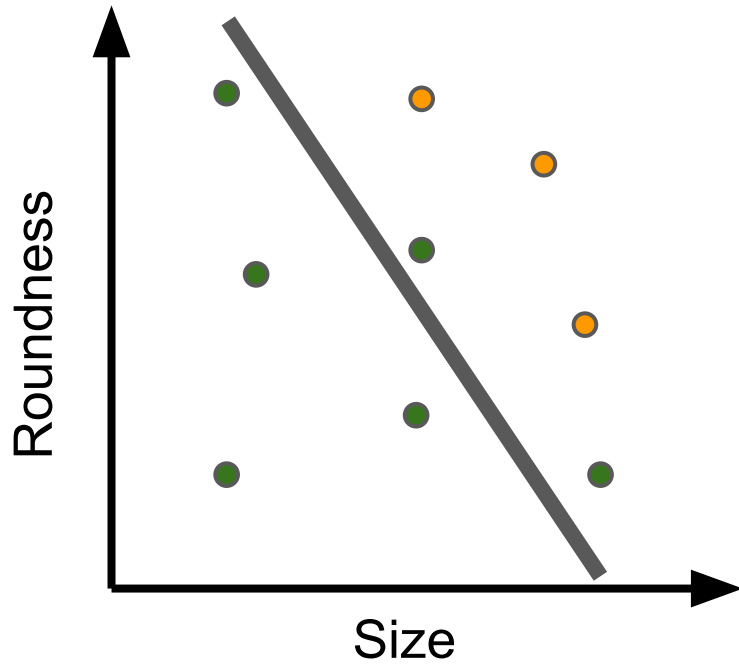
# What is learning?



# Desiderata

- Good performance on examples
- Good performance in future examples
- Computational costs
- Simplicity
- ...

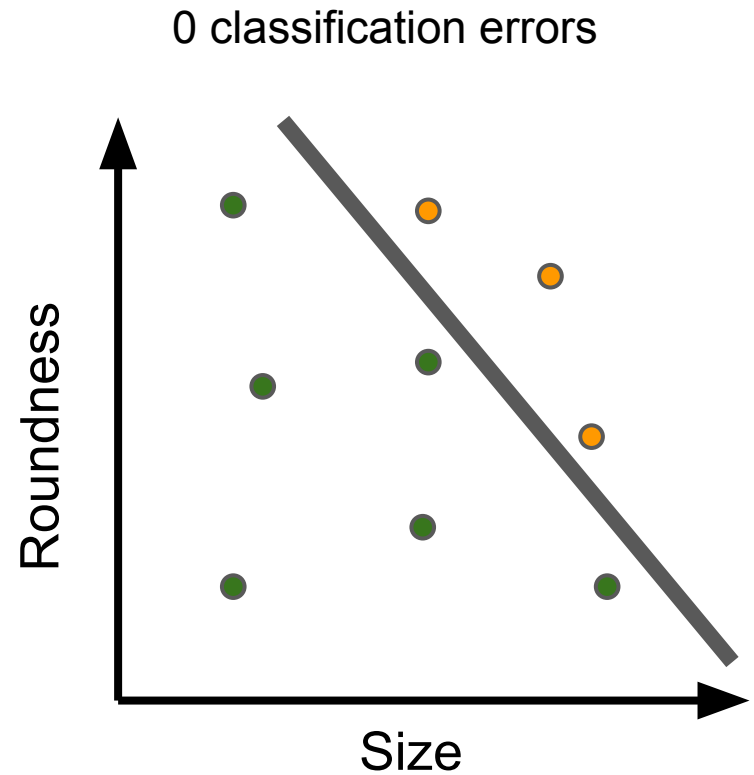
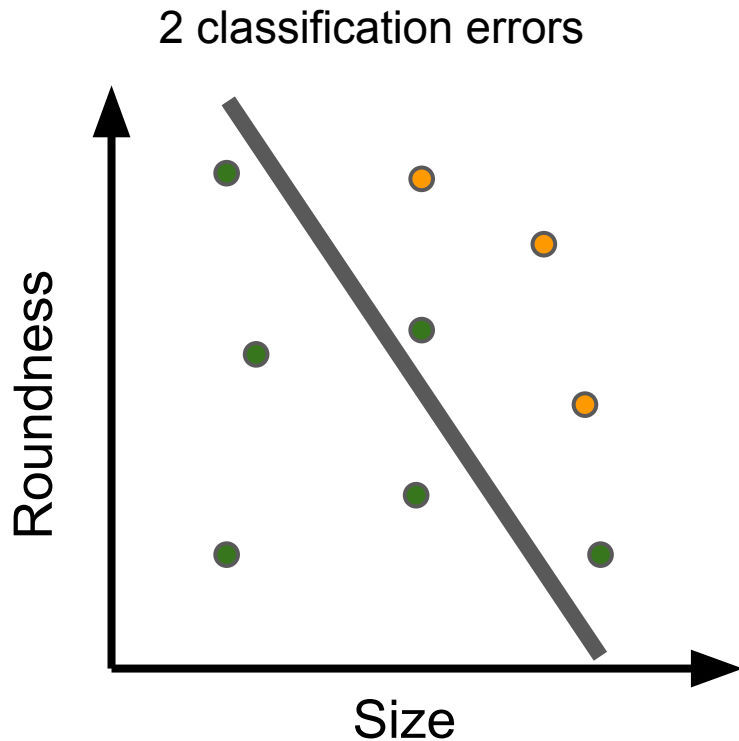
# Evaluating Performance



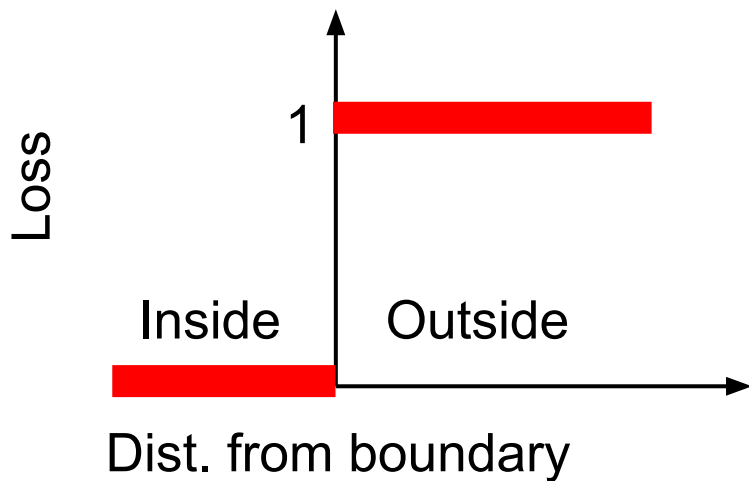
# Evaluating Performance

Loss function

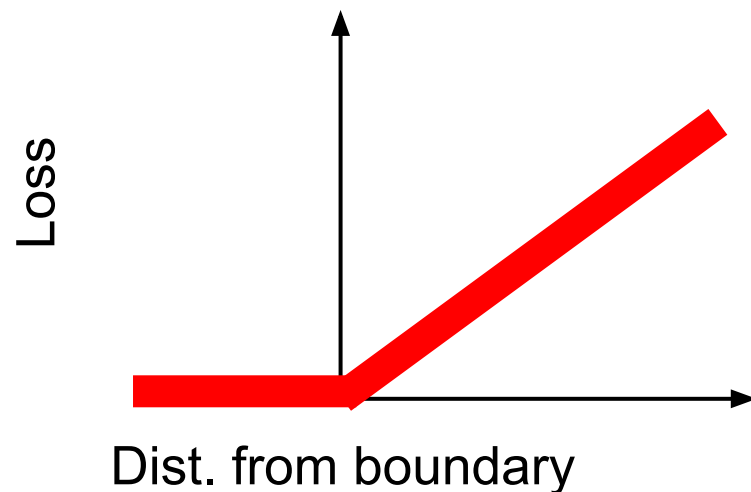
$$\text{loss}(\text{classifier}, \text{data}) = \dots$$



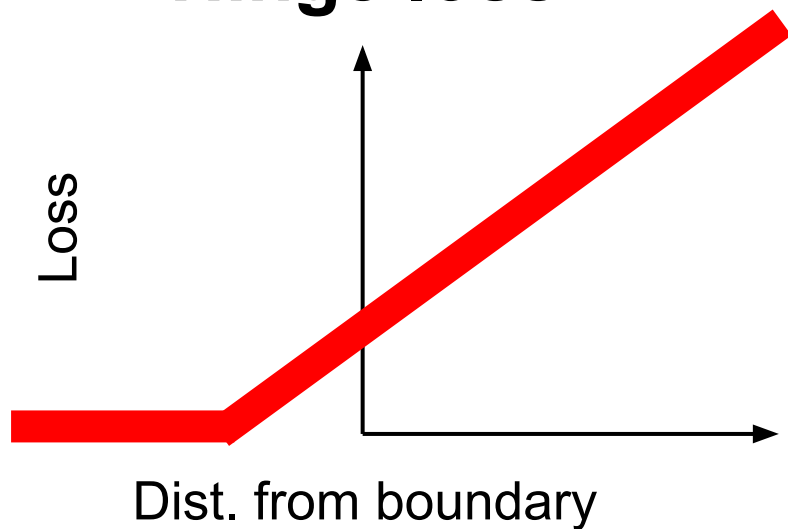
## 0/1 loss



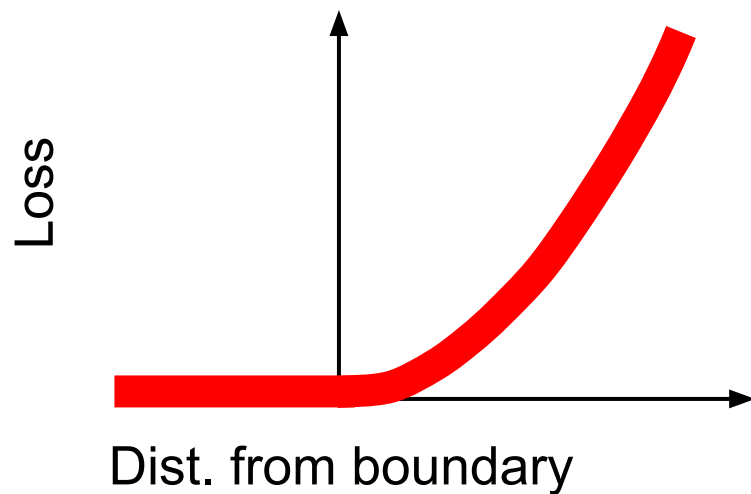
## Linear loss



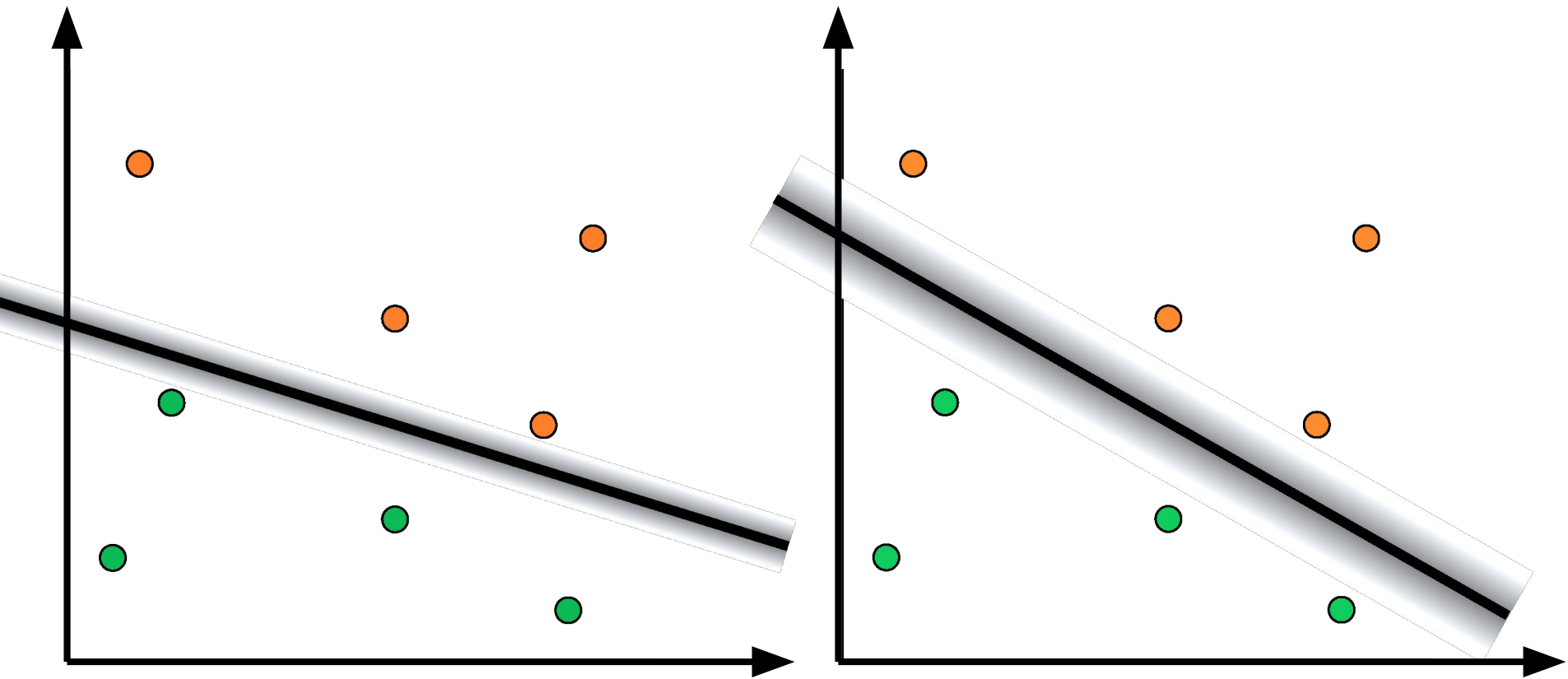
## Hinge loss



## Quadratic loss



# Beyond 0/1 loss





# Learning as optimization

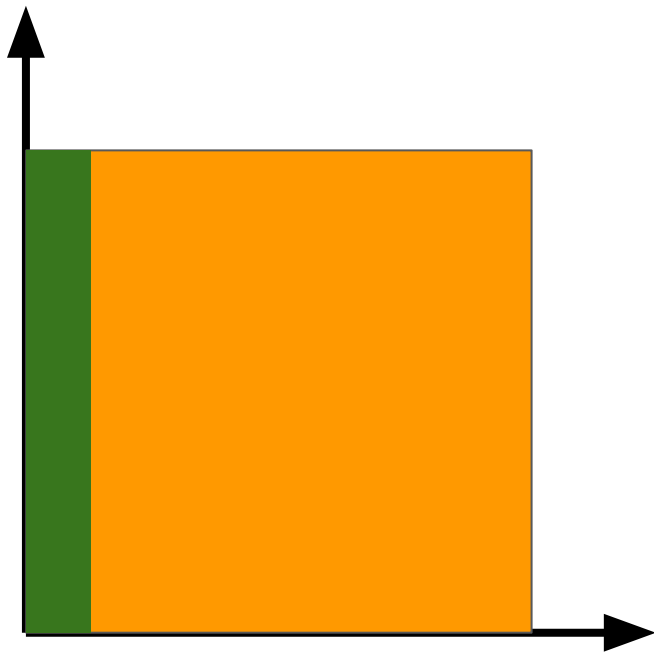
Given

- Data
- Family of classifiers
- Loss function

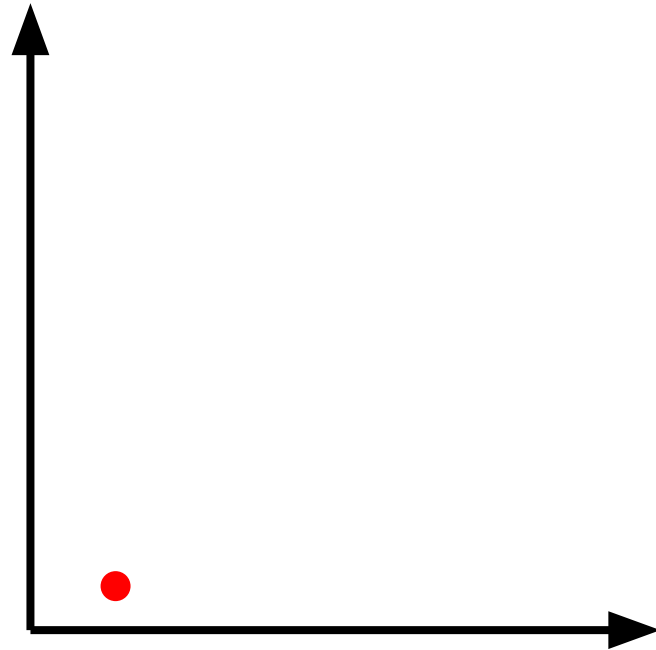
Find classifier with minimal loss

# Parameter space

Hypothesis space

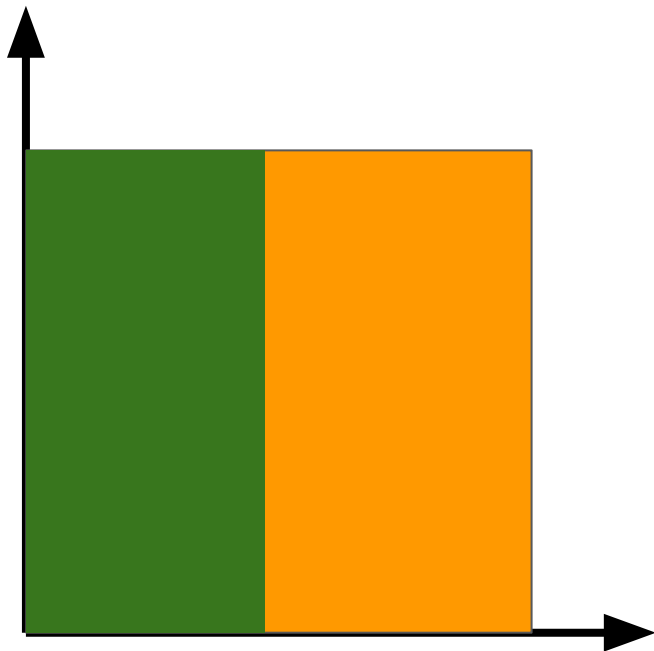


Parameter space

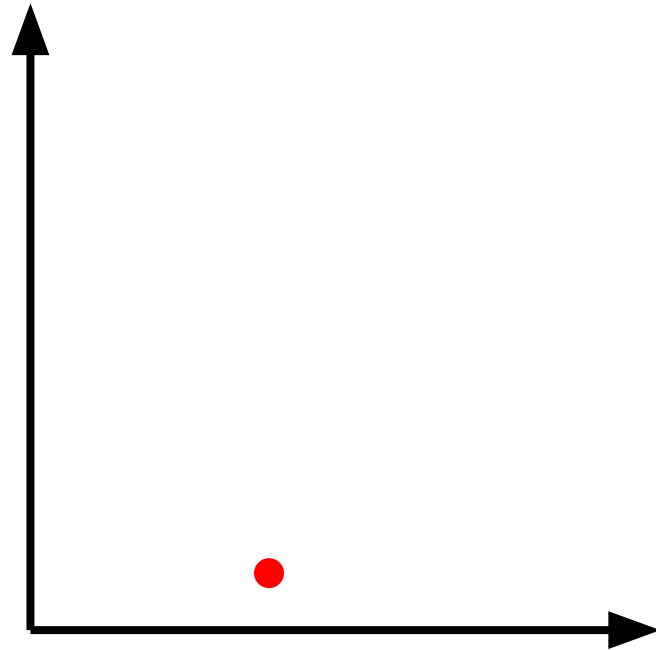


# Parameter space

Hypothesis space

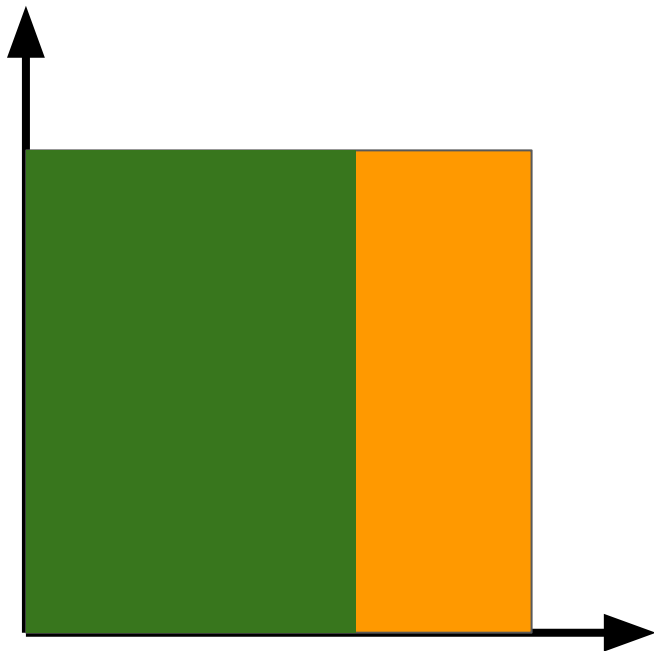


Parameter space

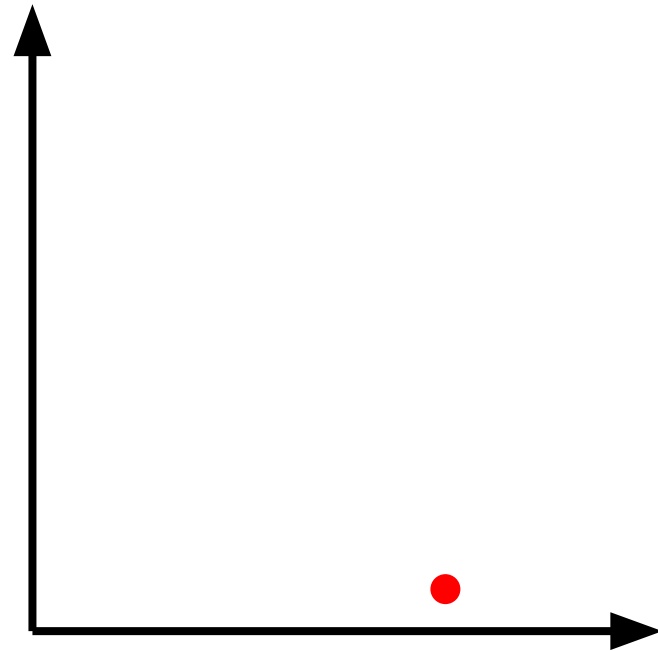


# Parameter space

Hypothesis space

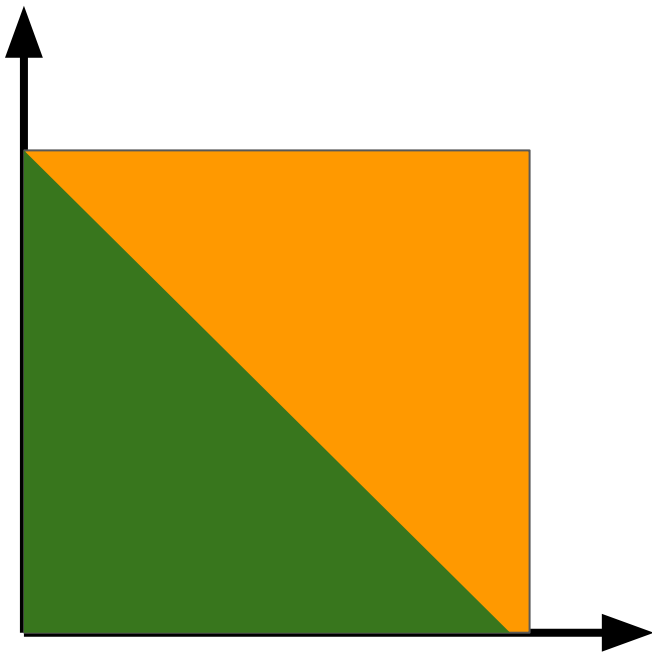


Parameter space

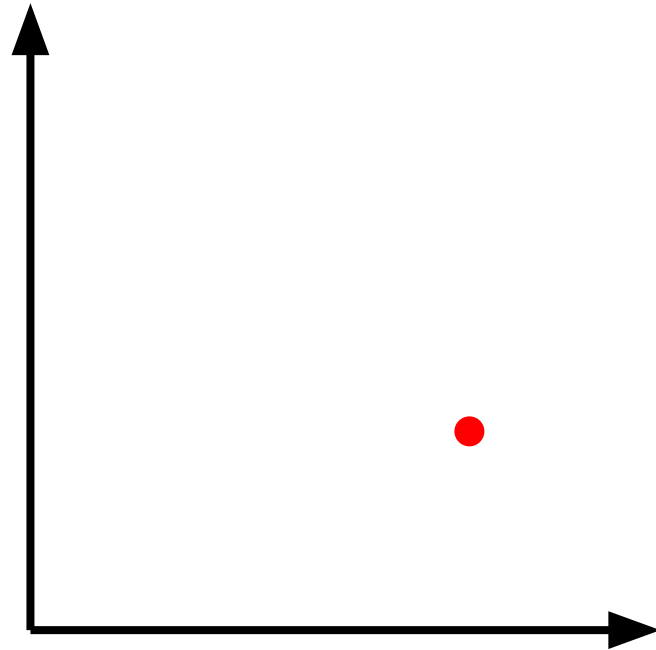


# Parameter space

Hypothesis space

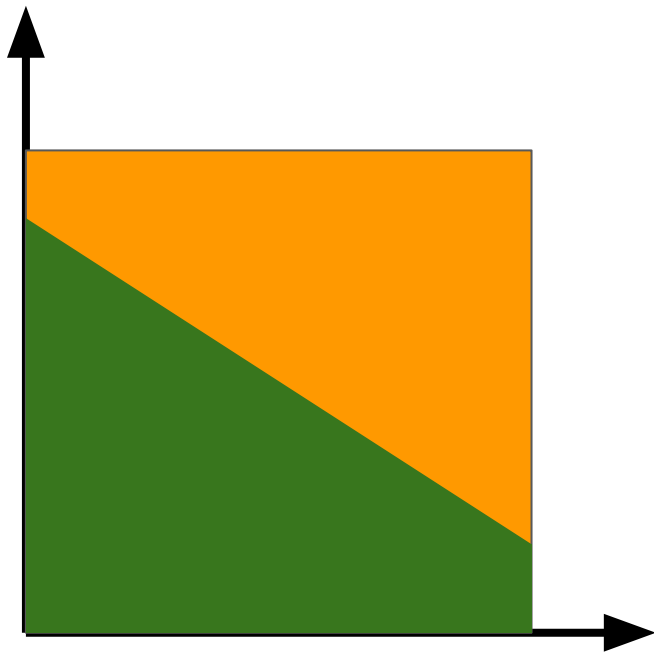


Parameter space

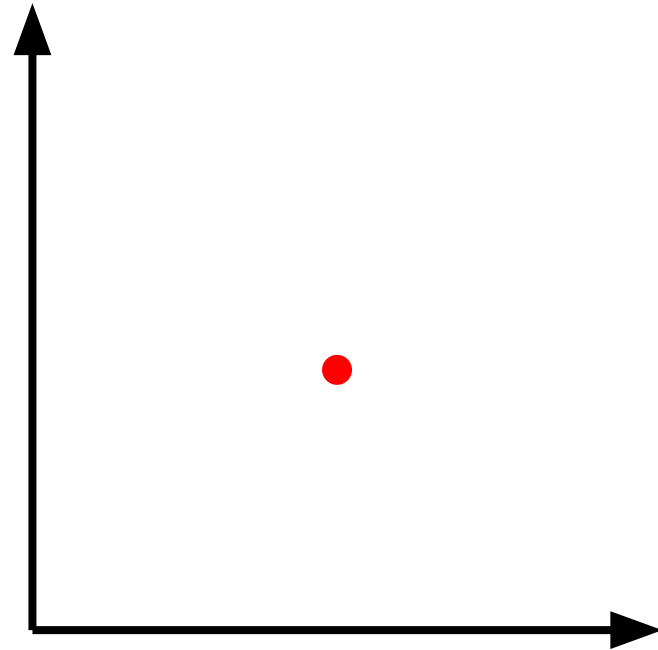


# Parameter space

Hypothesis space

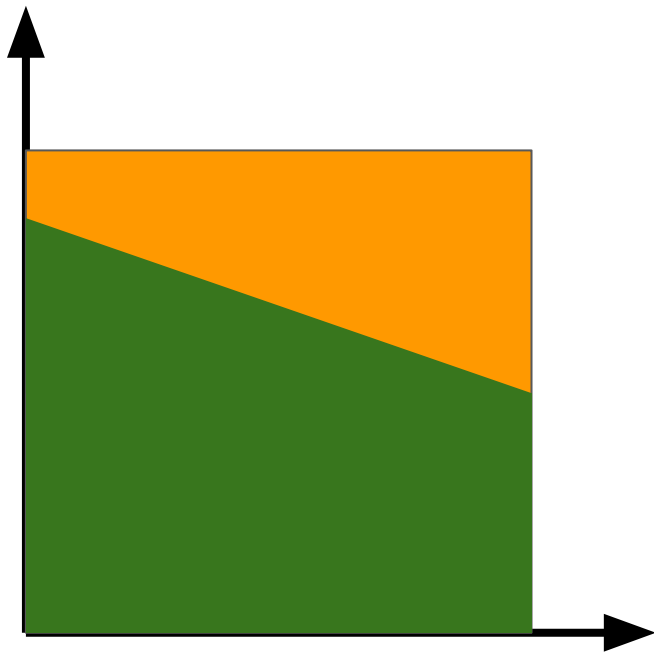


Parameter space

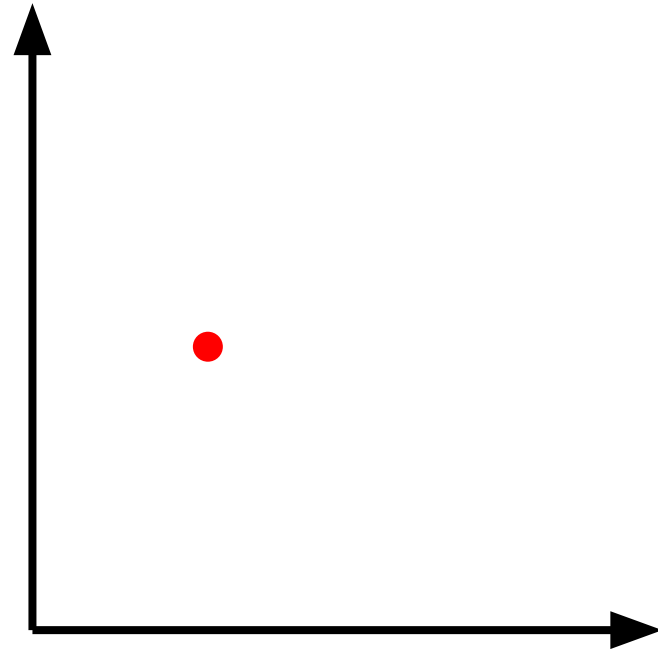


# Parameter space

Hypothesis space

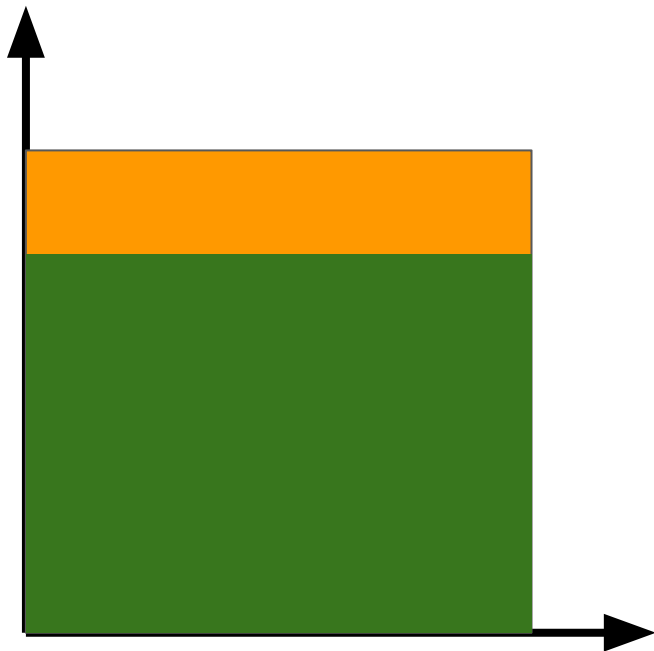


Parameter space

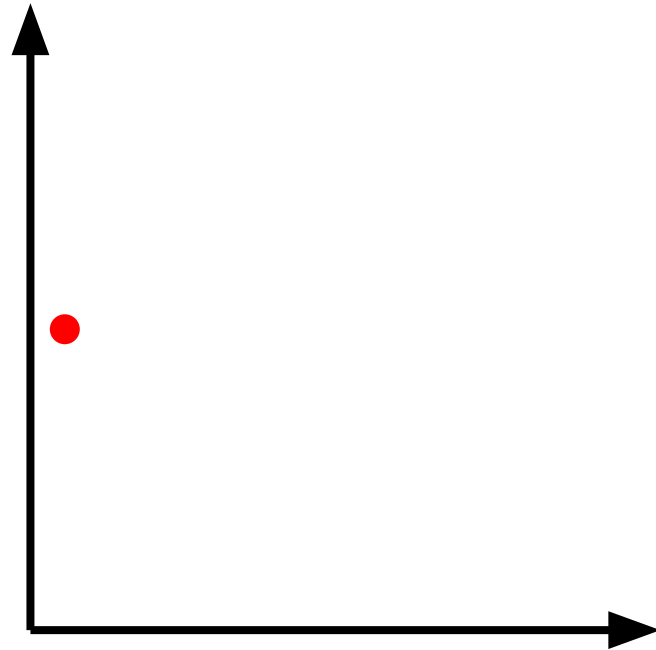


# Parameter space

Hypothesis space



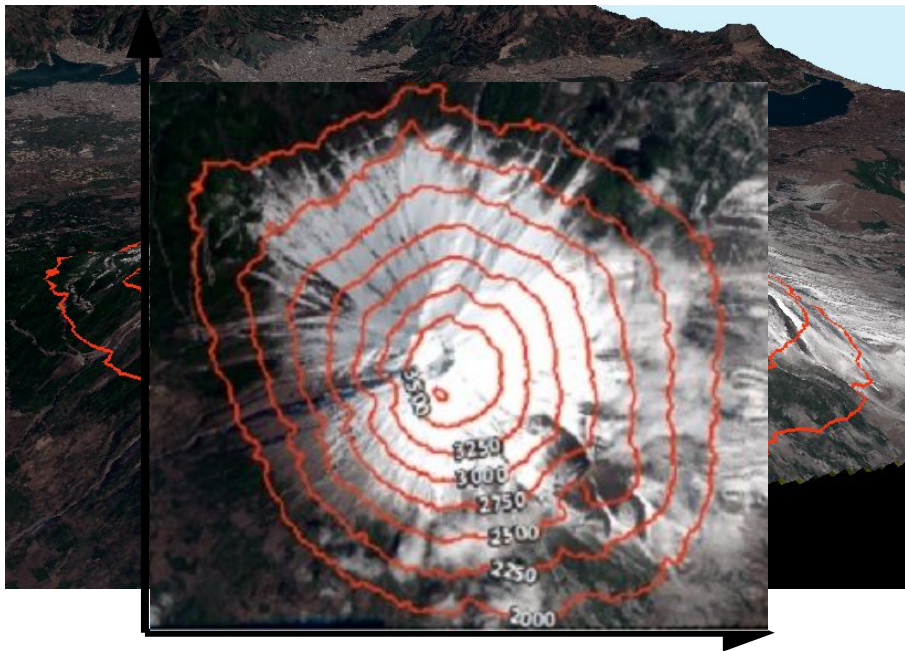
Parameter space



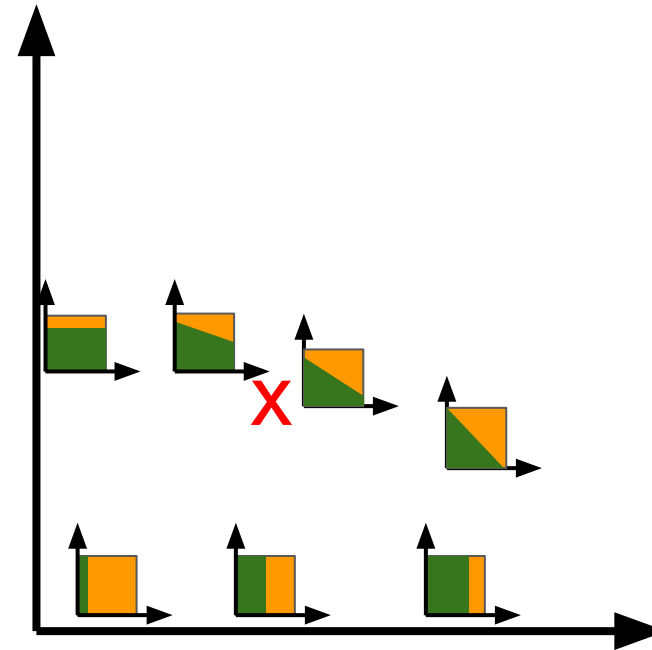


# Parameter space optimization

Optimum search



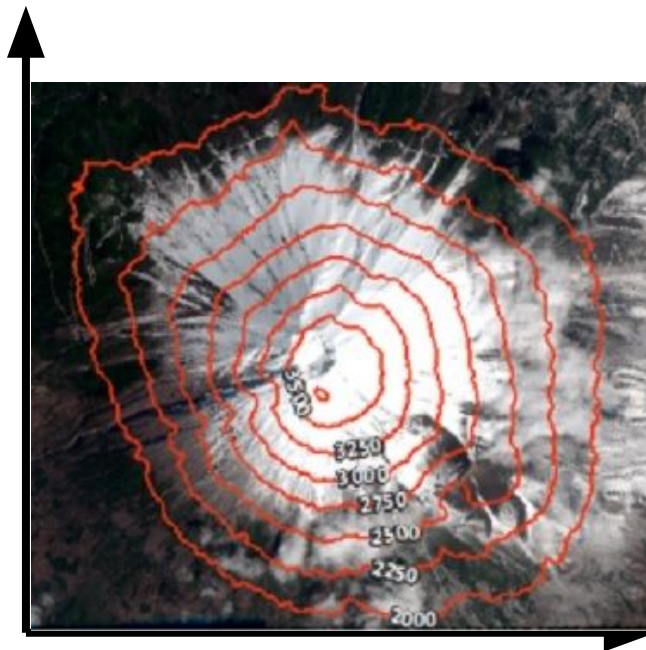
Parameter space



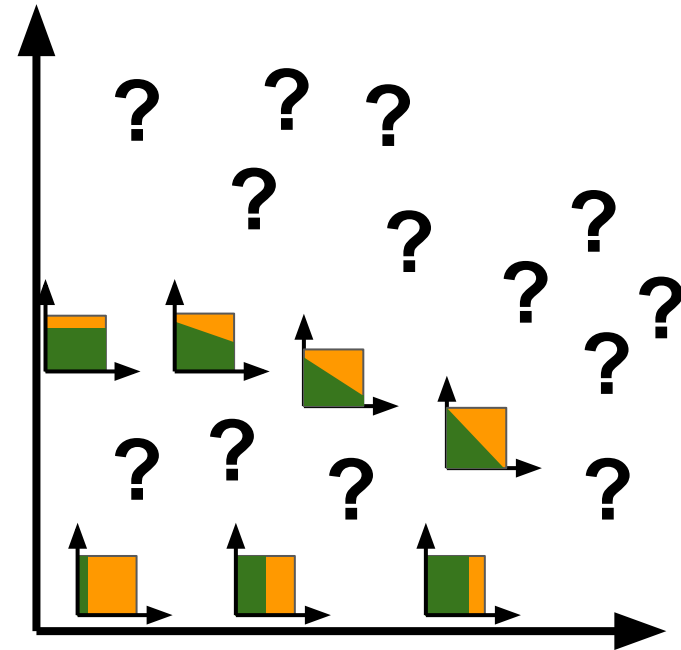
# Optimization 101

Dealing with partial knowledge

Optimum search

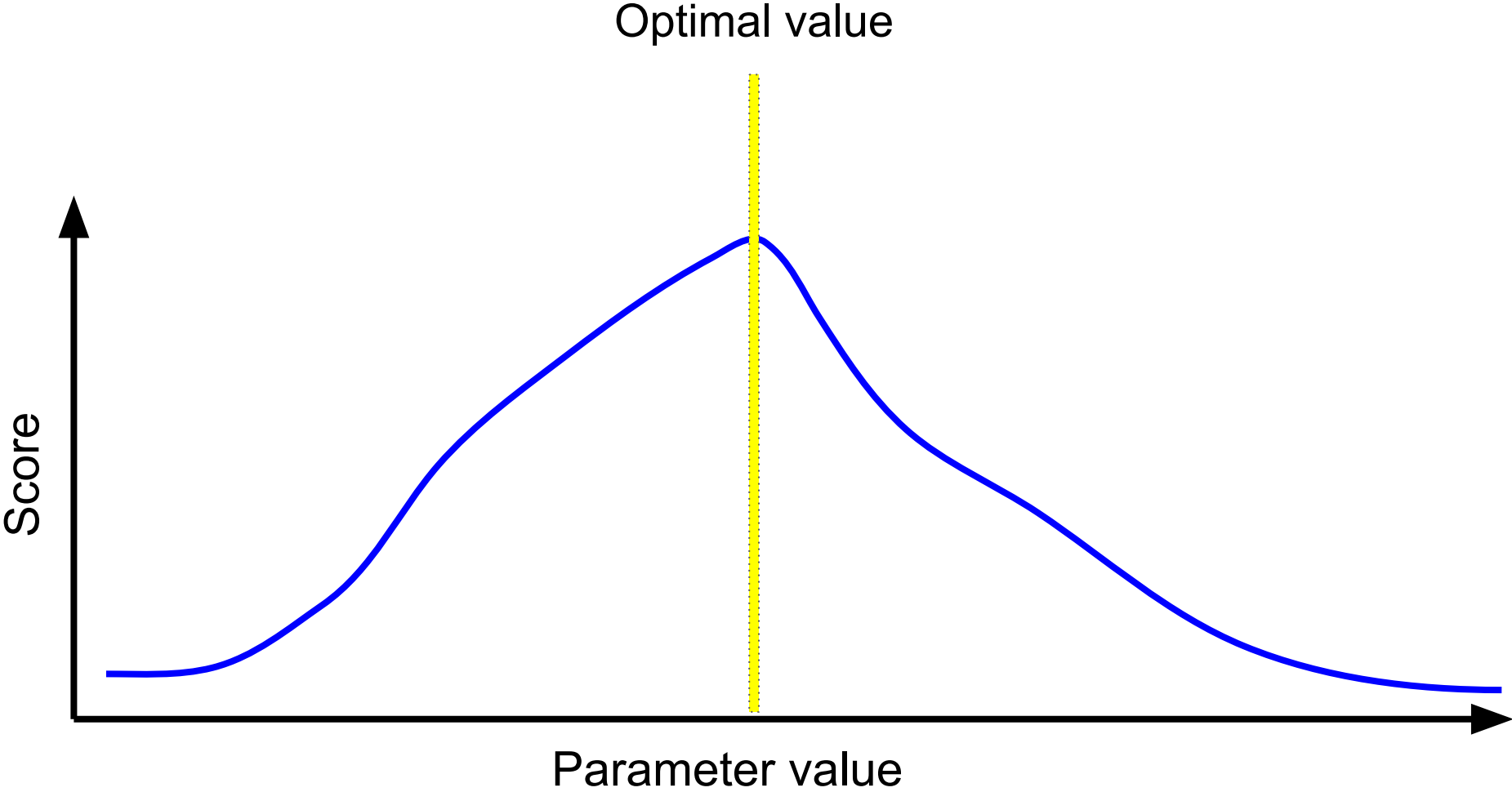


Parameter space



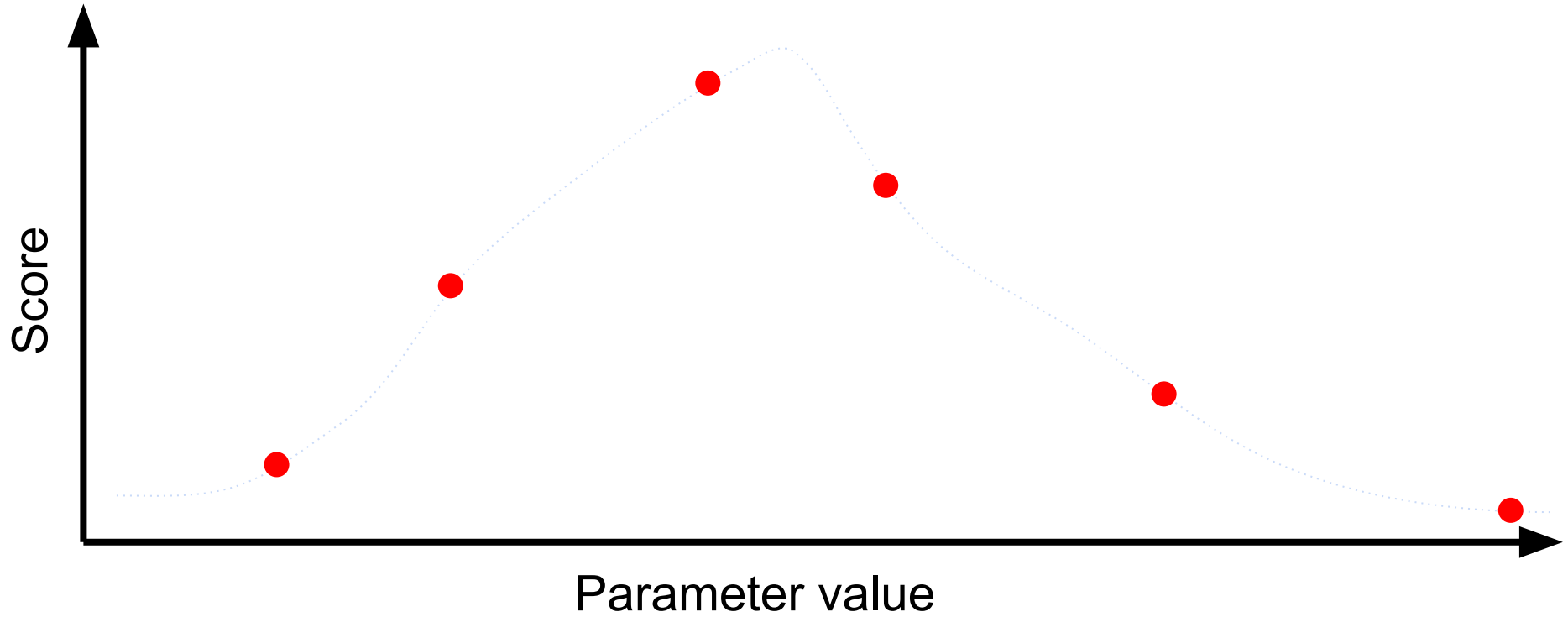
# Optimization 101

Line search



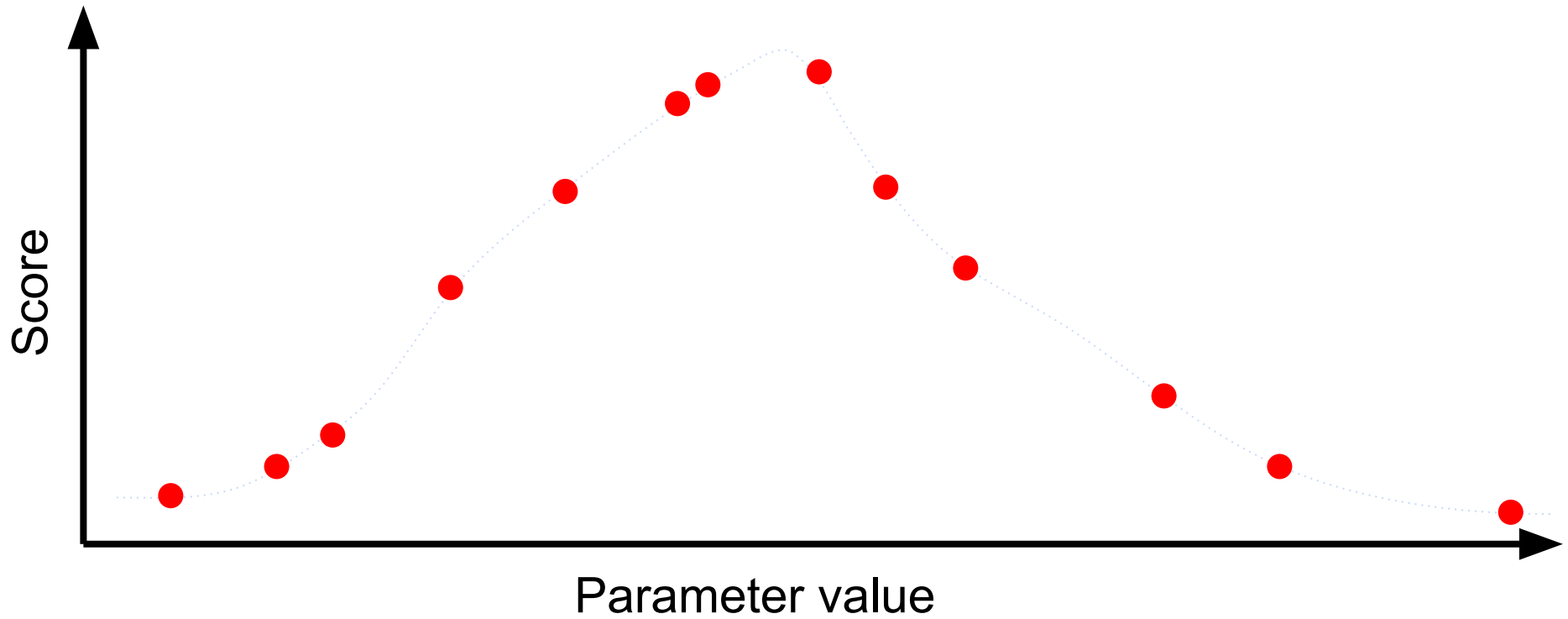
# Optimization 101

Line search



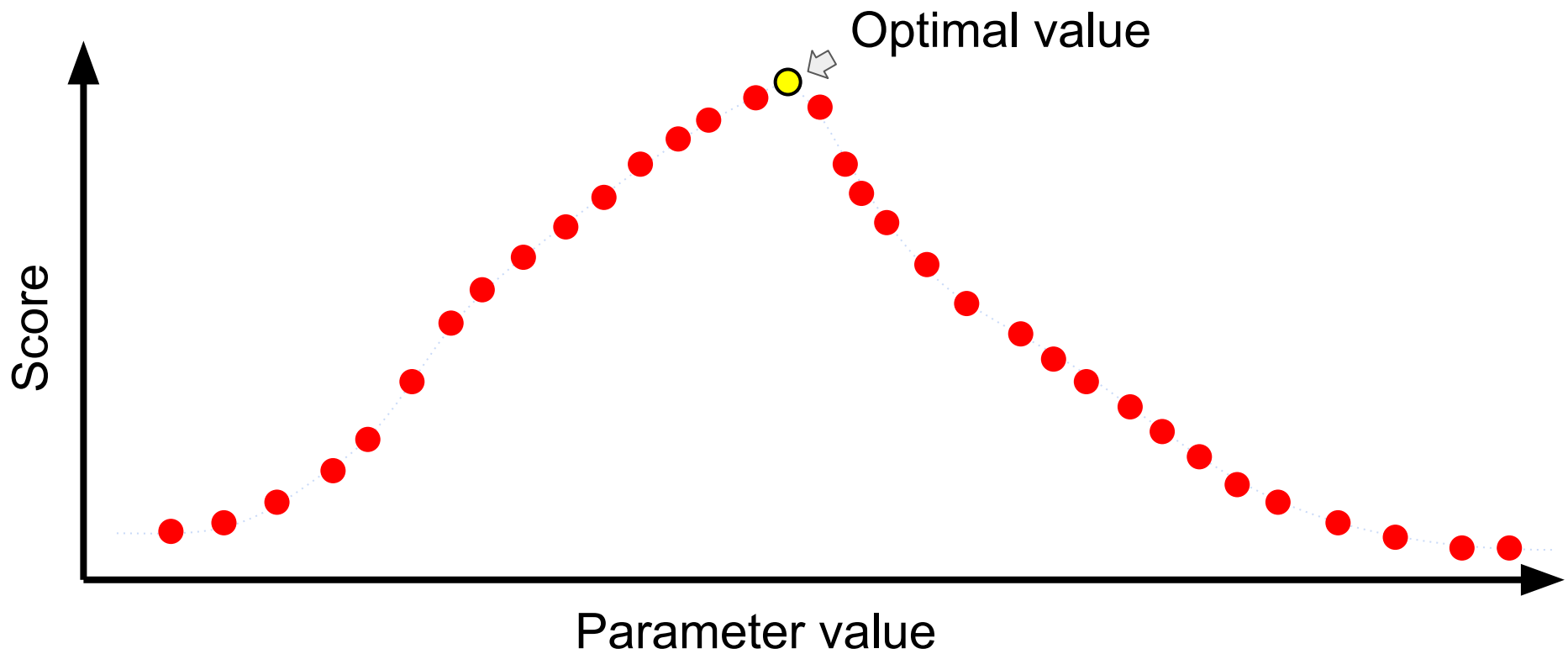
# Optimization 101

Line search



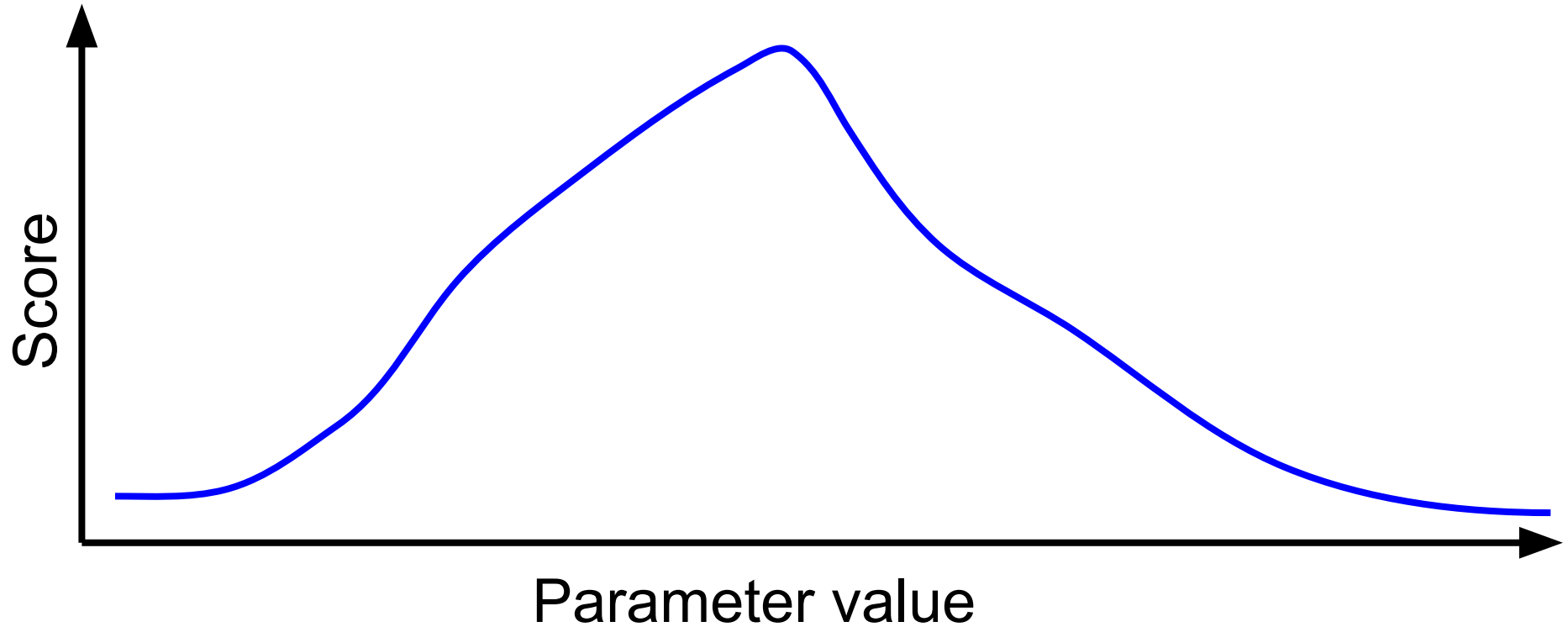
# Optimization 101

Line search via brute force enumeration



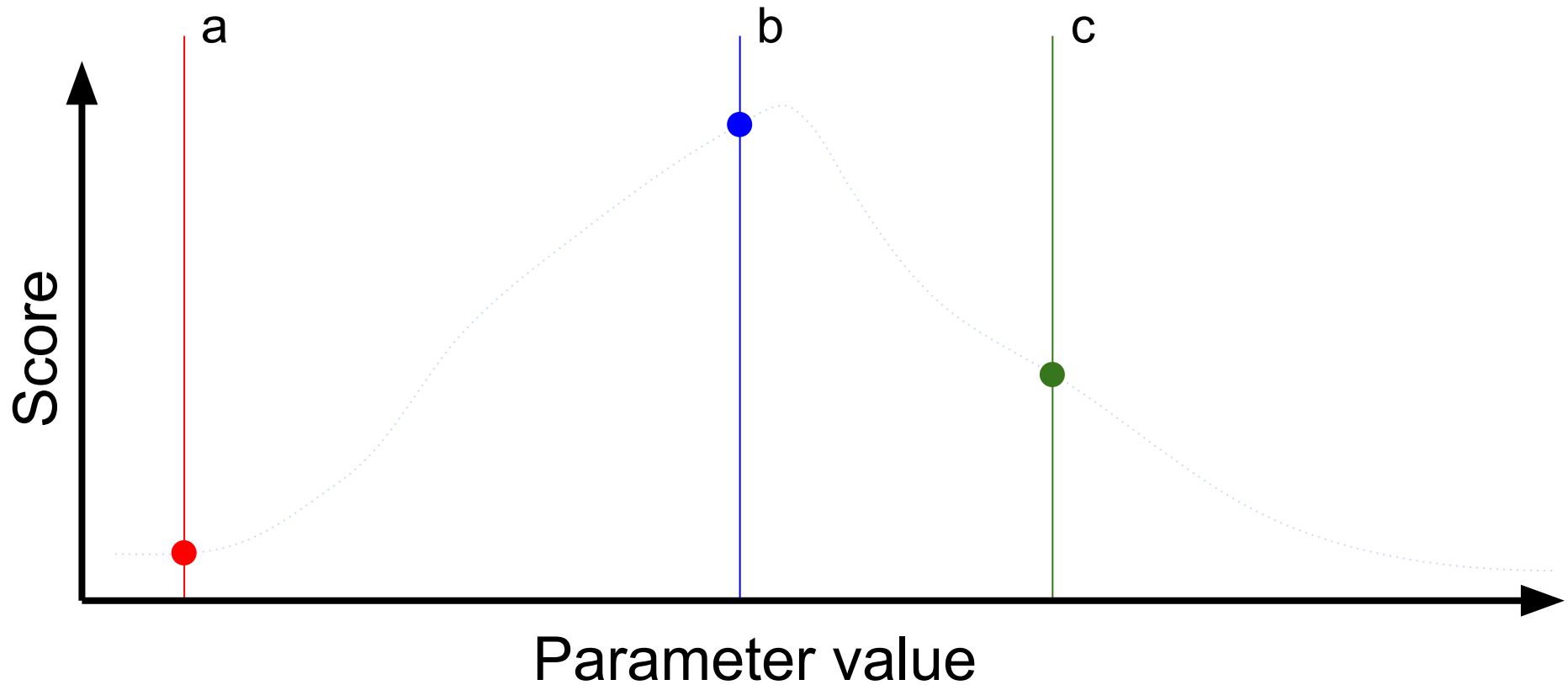
# Optimization 101

Line search



# Optimization 101

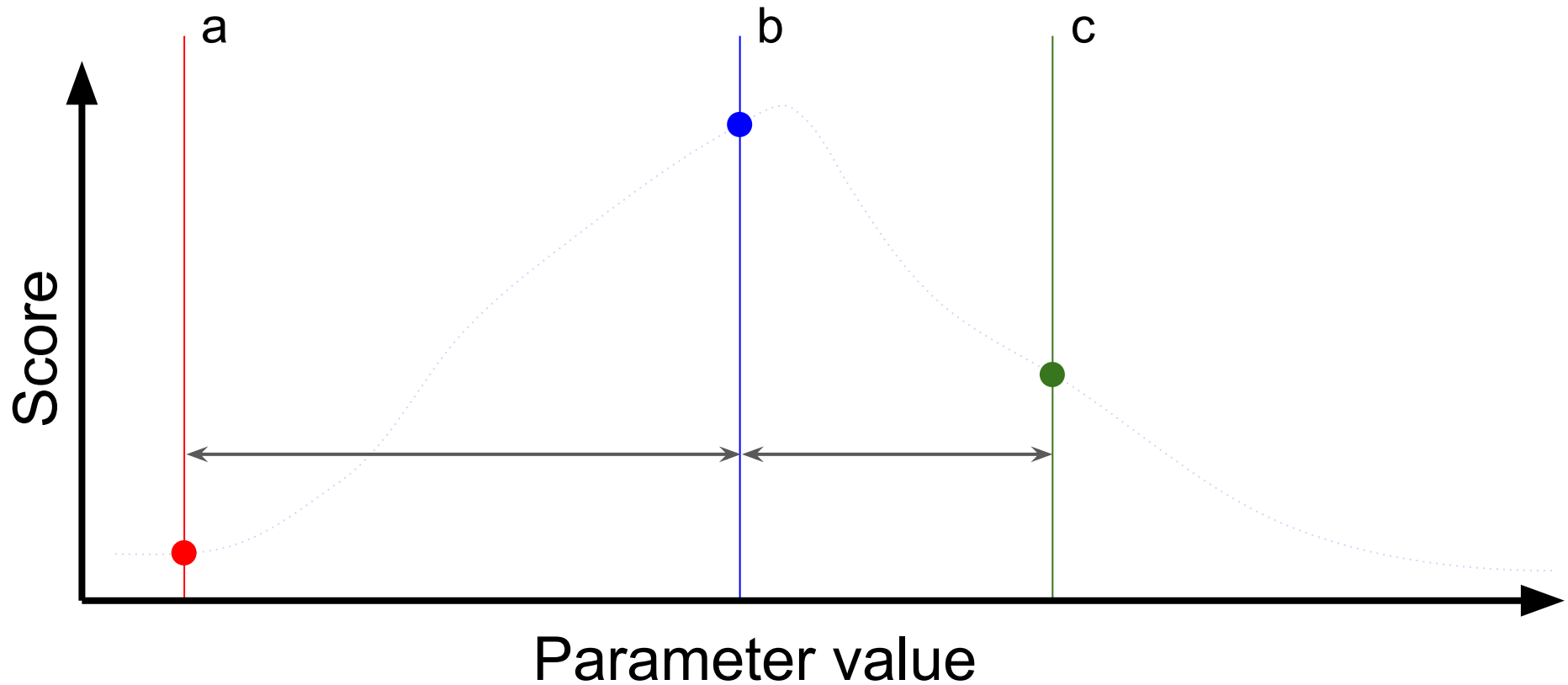
Line search using bisection





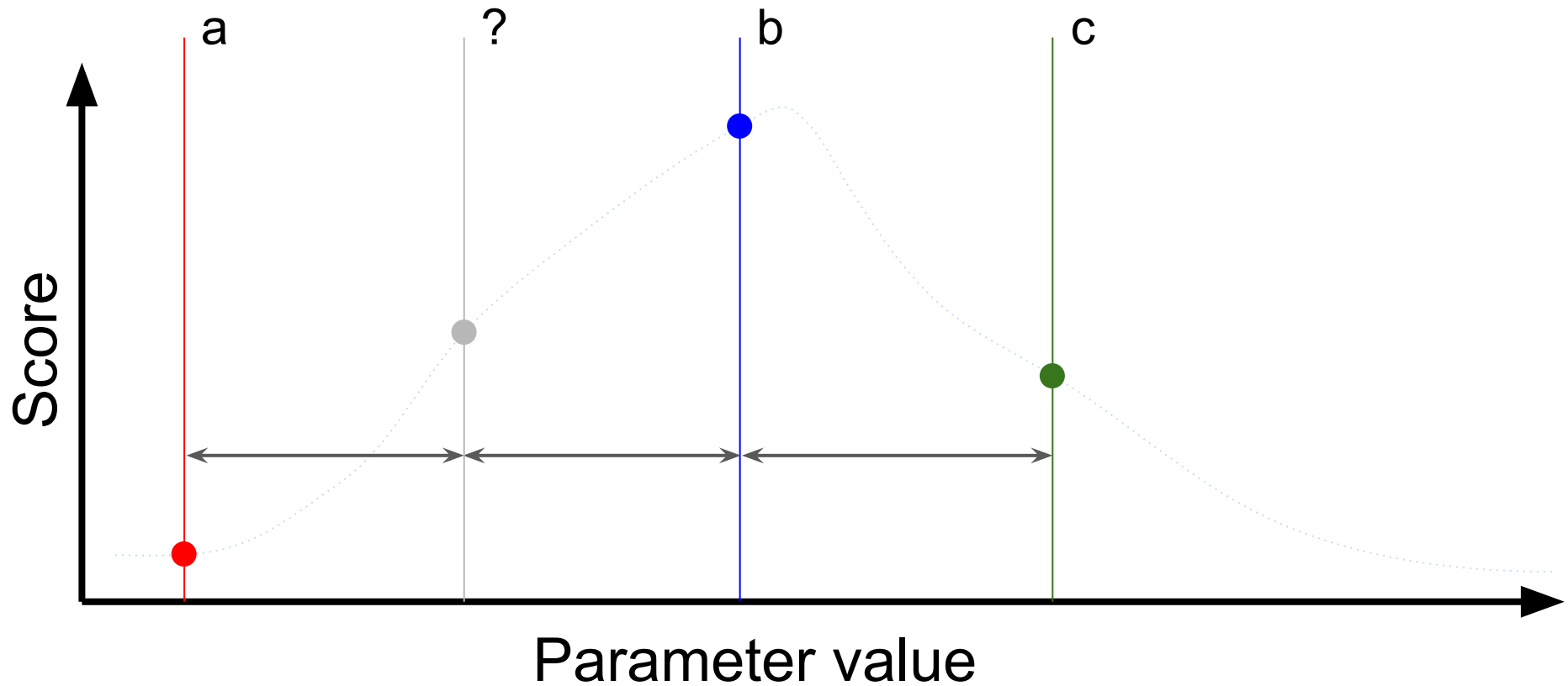
# Optimization 101

Line search using bisection



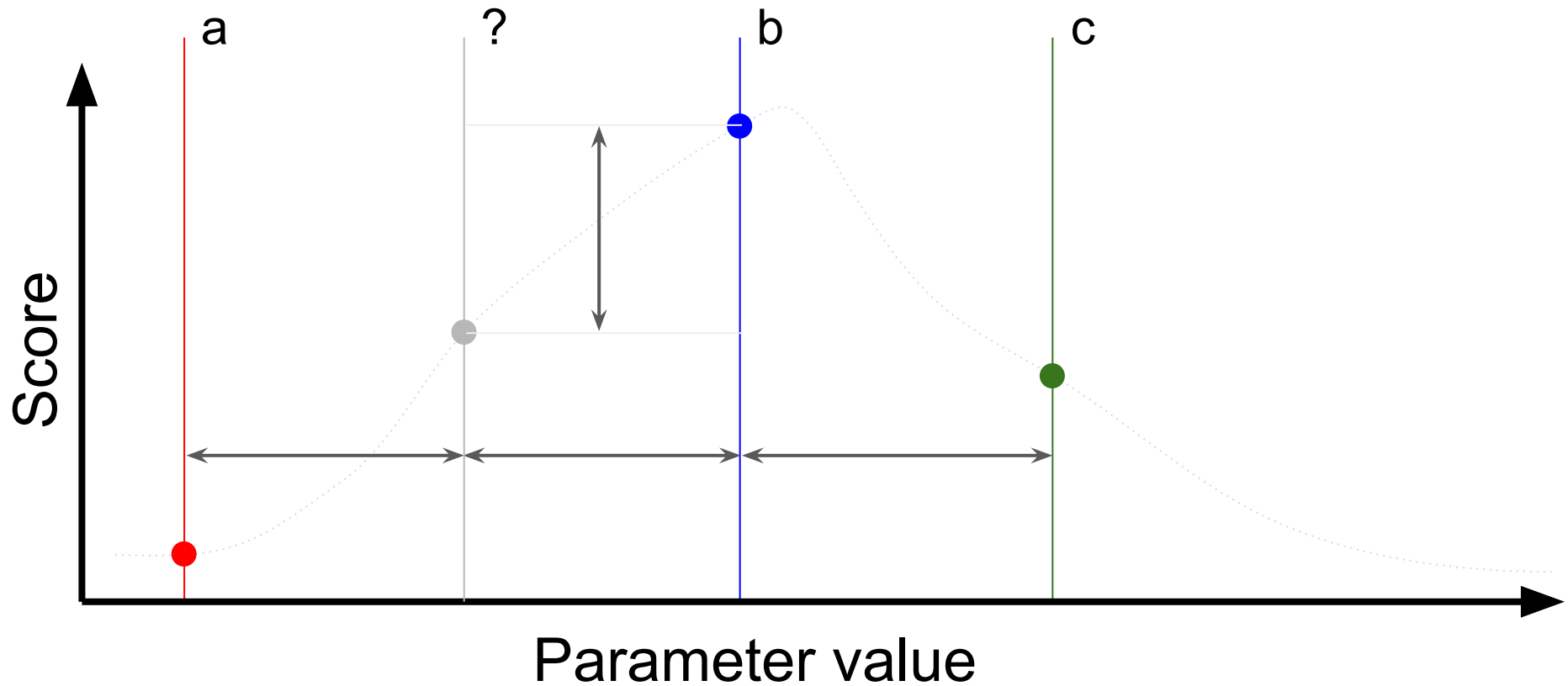
# Optimization 101

Line search using bisection



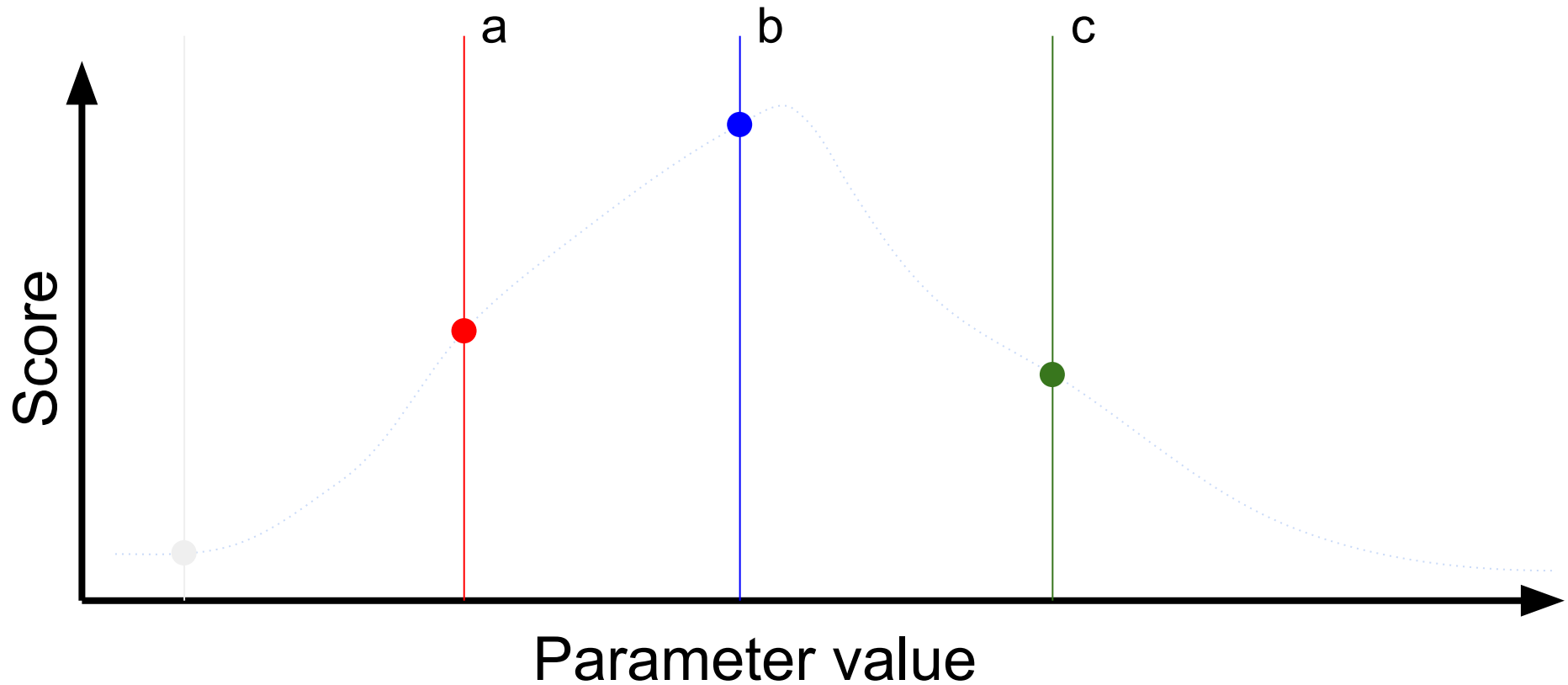
# Optimization 101

Line search using bisection



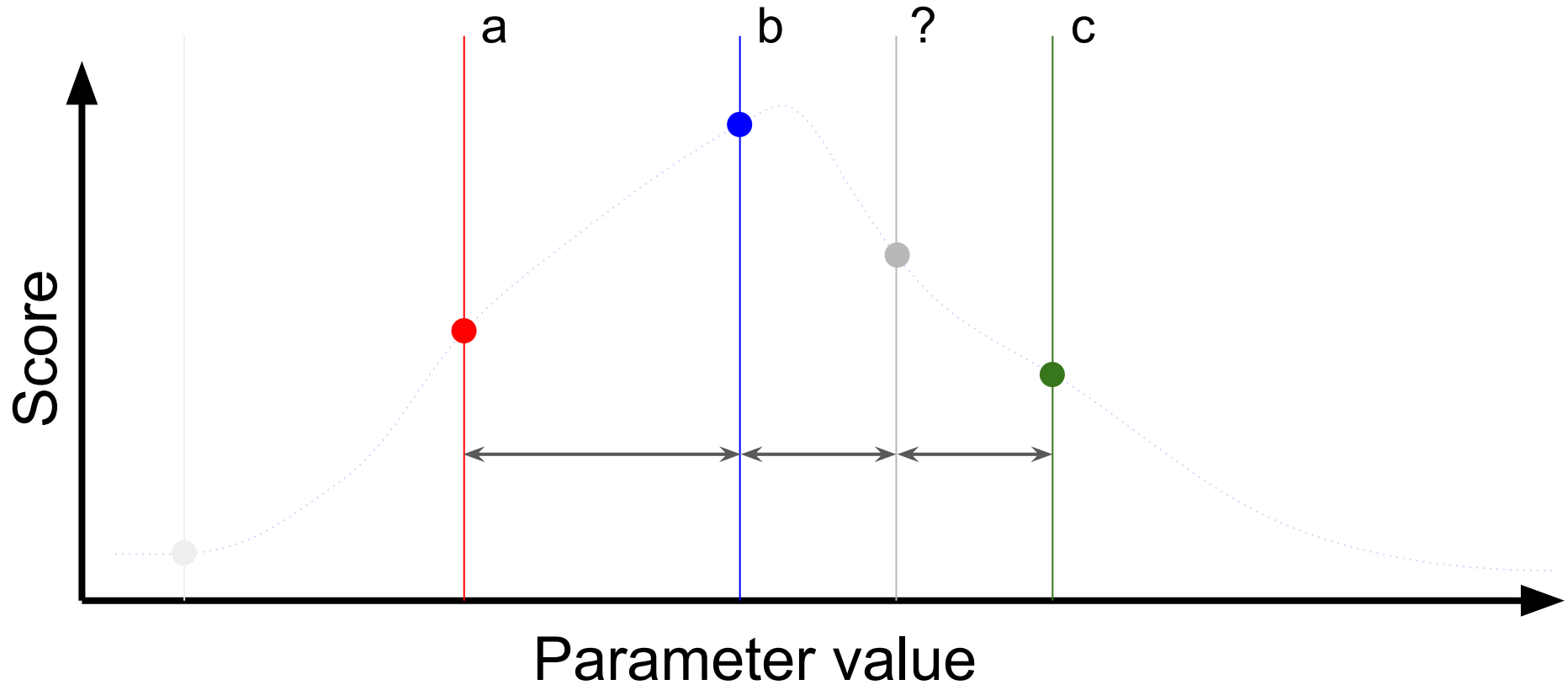
# Optimization 101

Line search using bisection



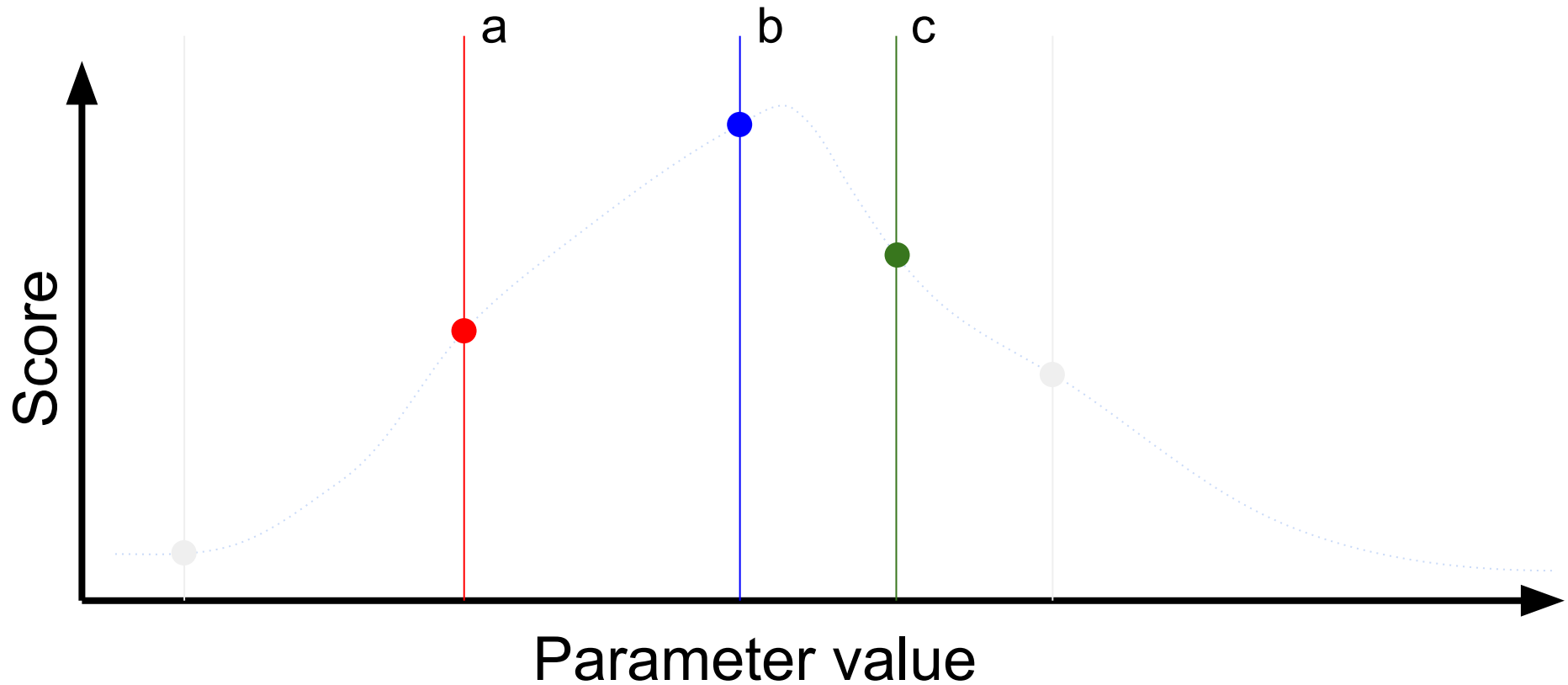
# Optimization 101

Line search using bisection



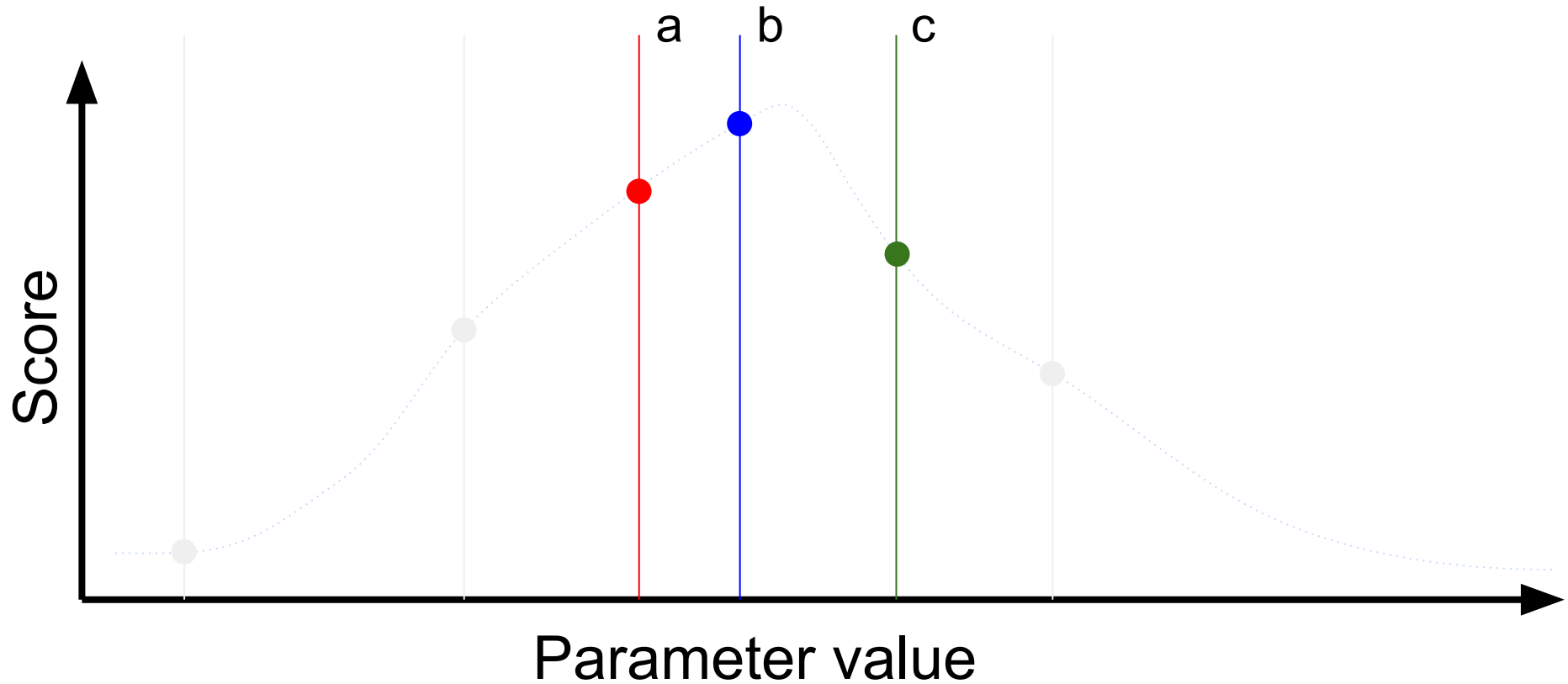
# Optimization 101

Line search using bisection



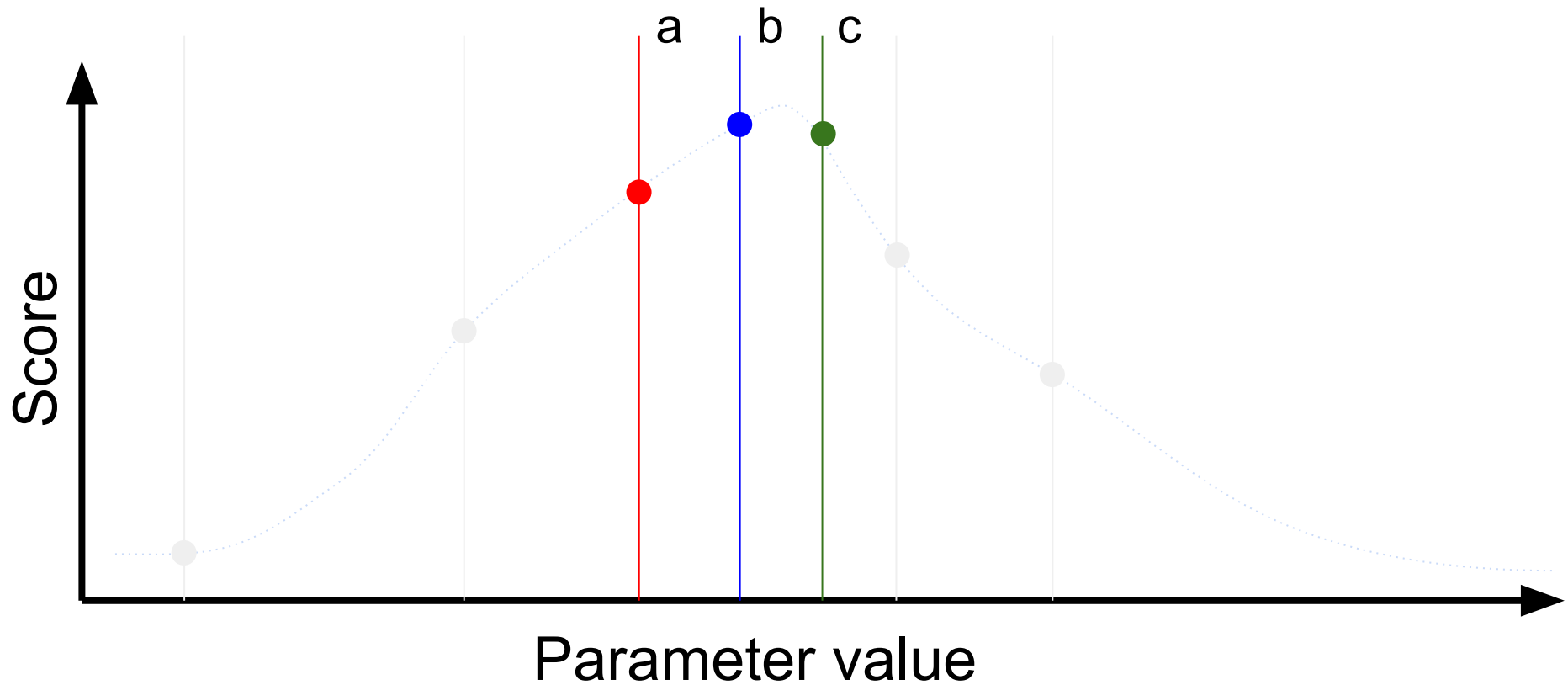
# Optimization 101

Line search using bisection



# Optimization 101

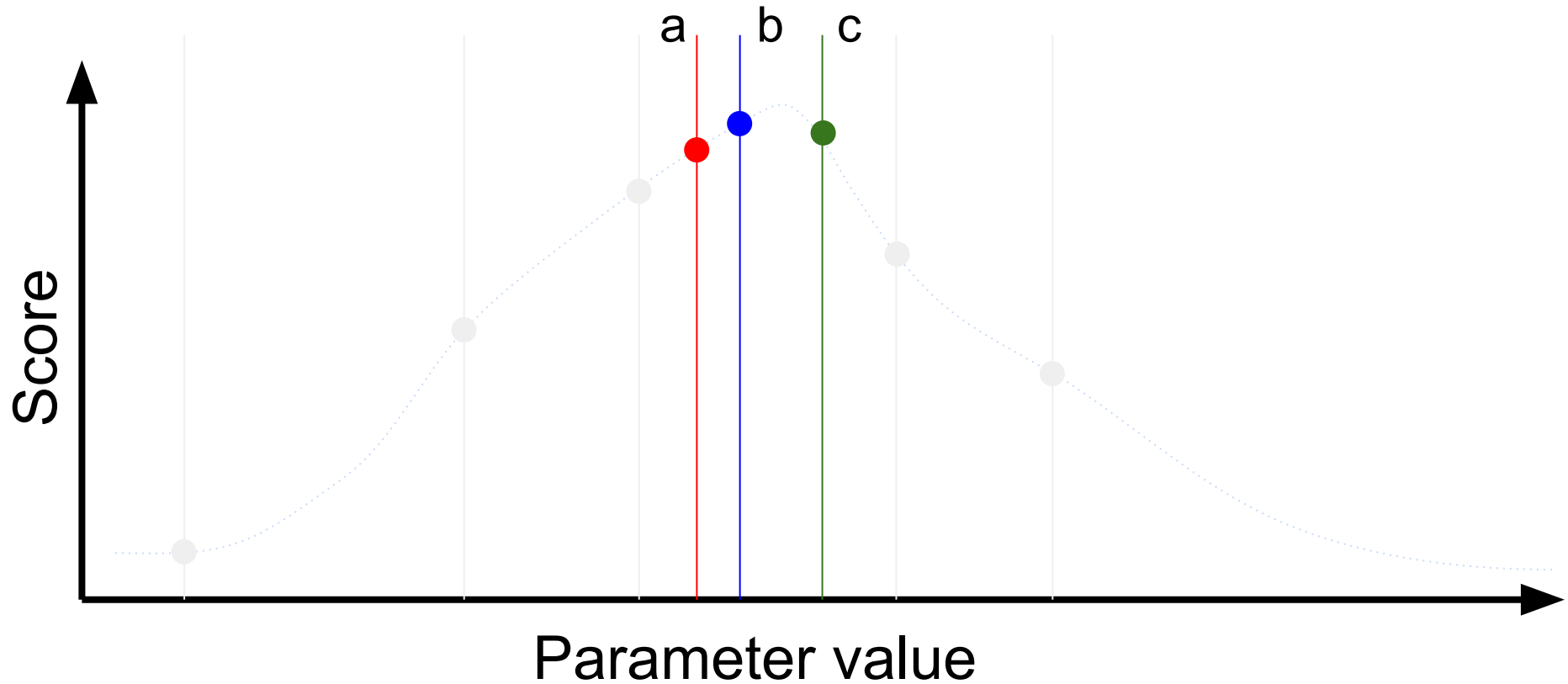
Line search using bisection





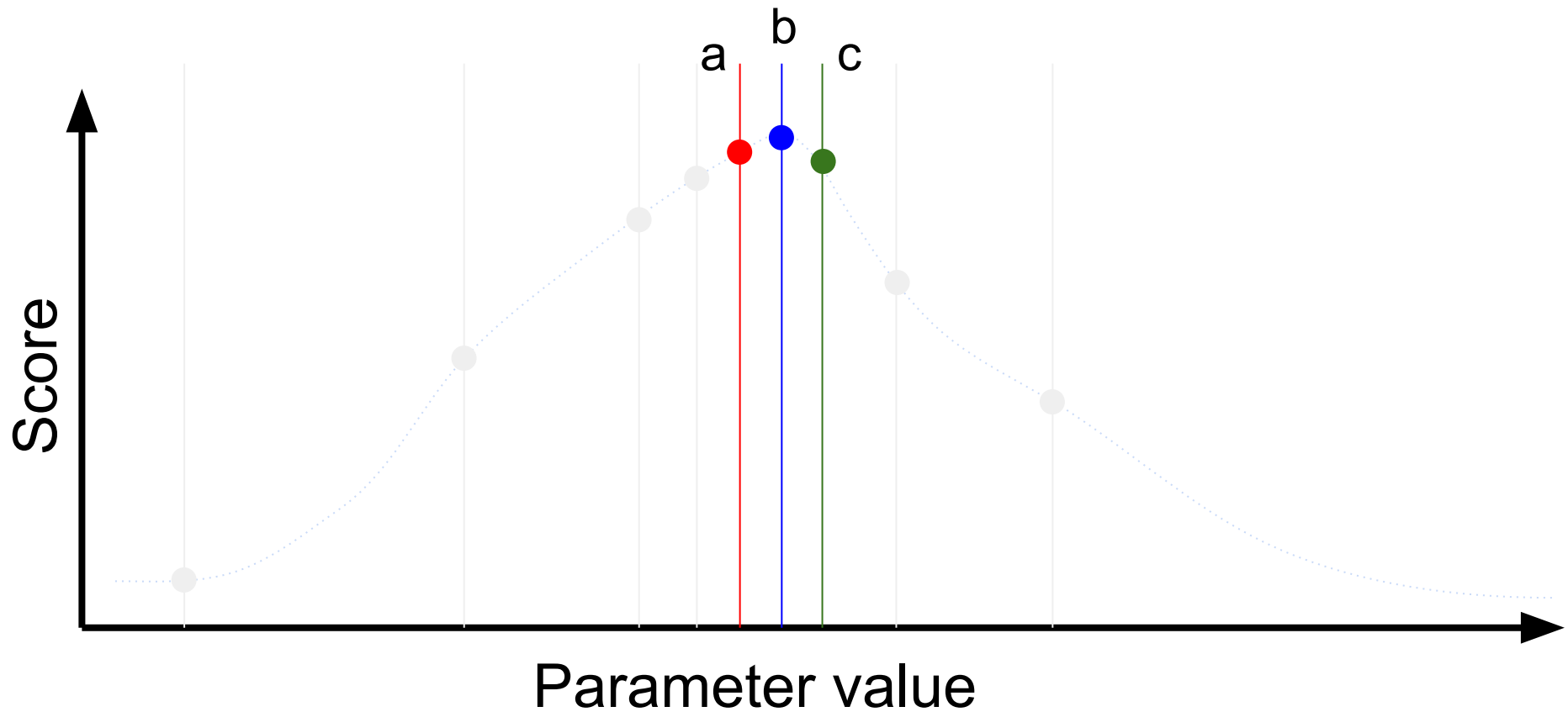
# Optimization 101

Line search using bisection



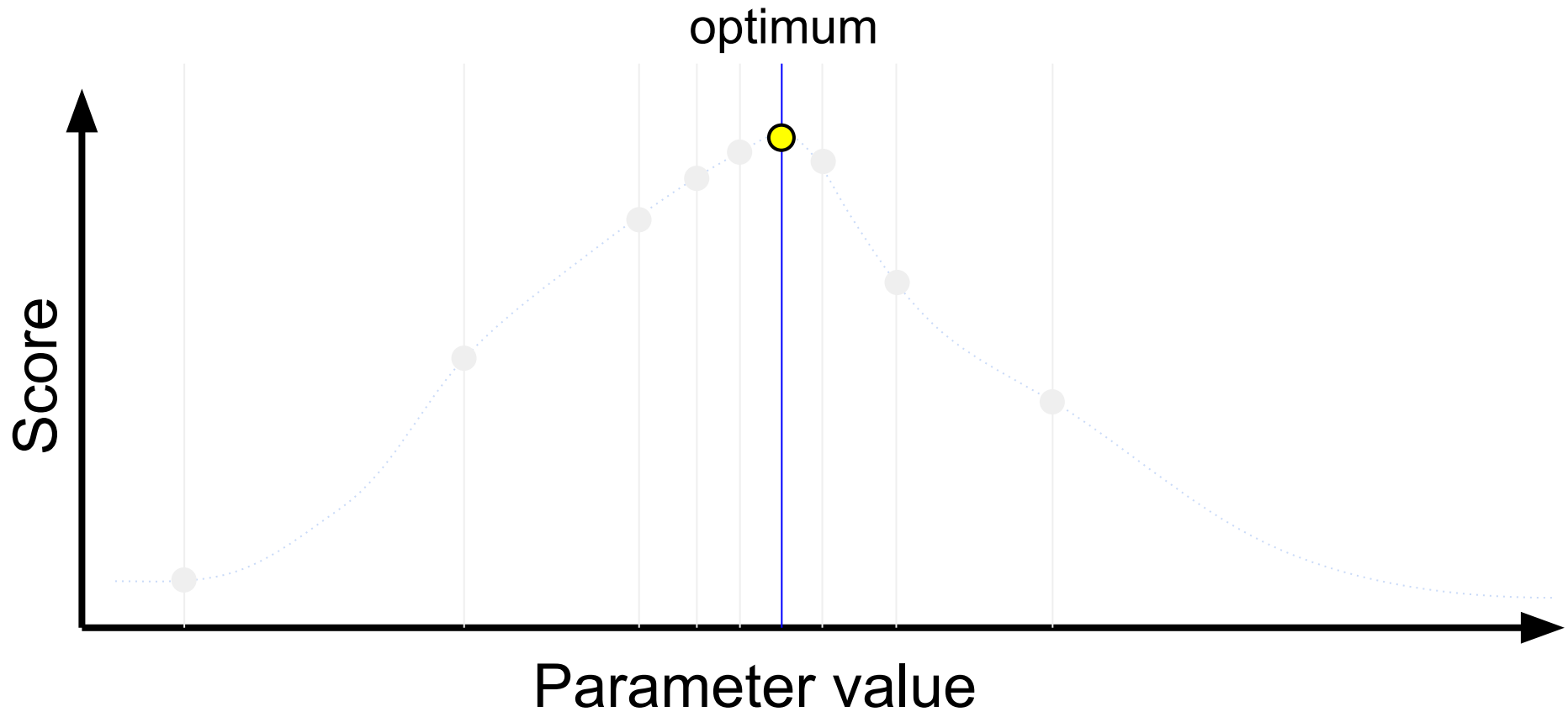
# Optimization 101

Line search using bisection



# Optimization 101

Line search using bisection



# Optimization 101

Higher order spaces



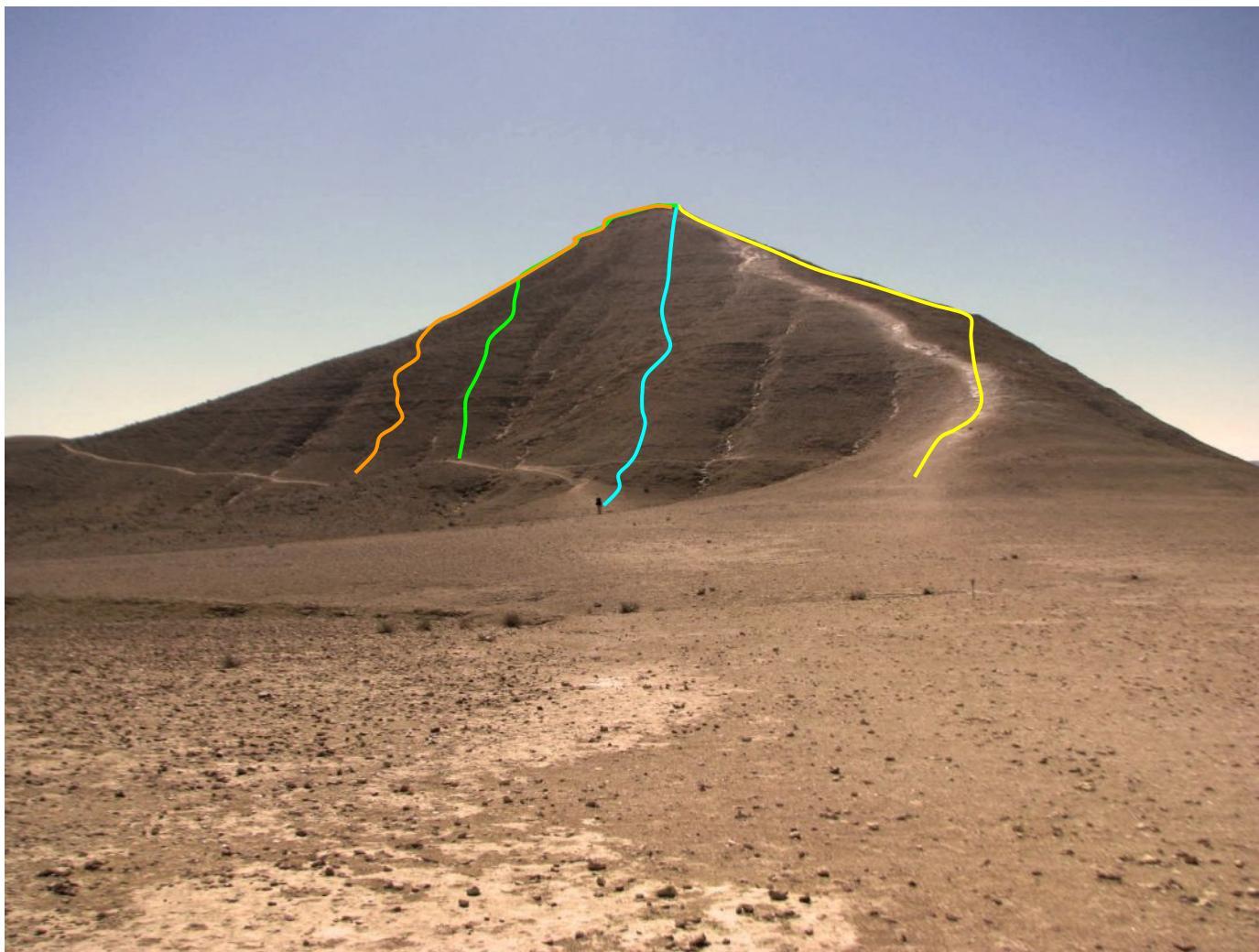
# Optimization 101

Steepest descent



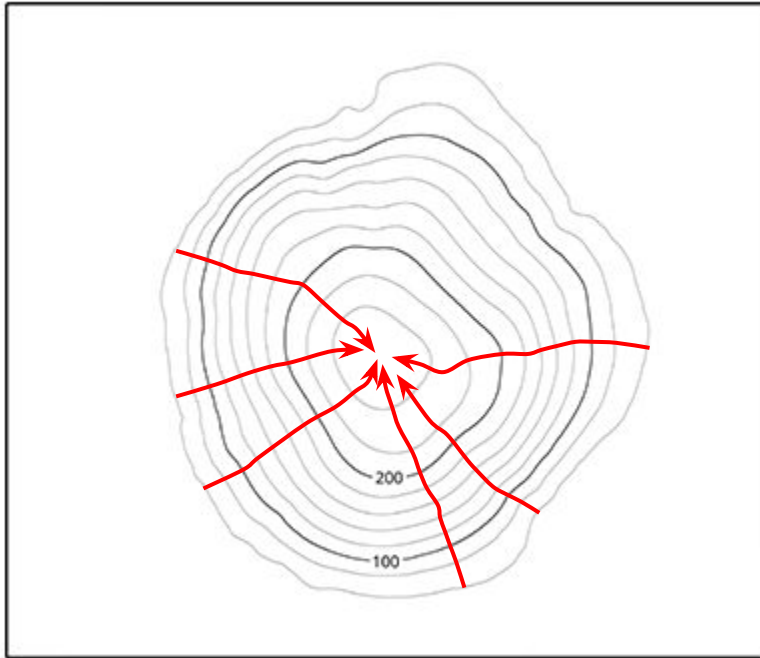
# Optimization 101

Steepest ascent



# Optimization 101

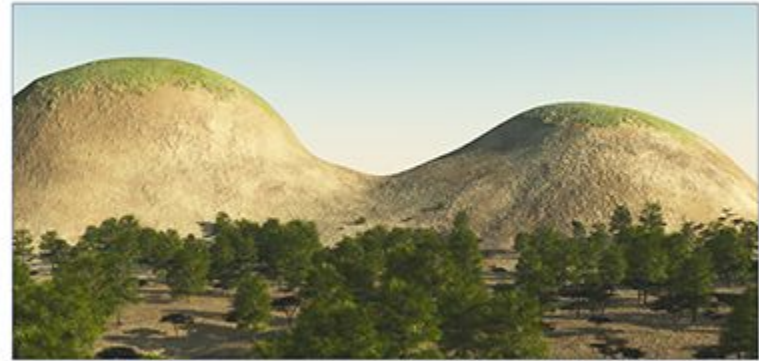
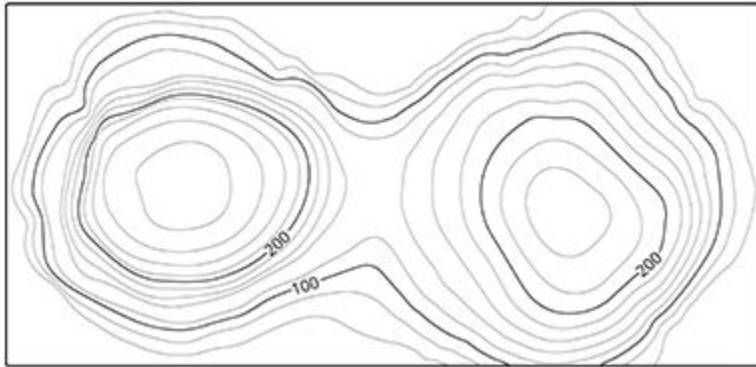
## Gradient ascent



$$\nabla(f) = \frac{\partial f}{\partial x} e_1 + \frac{\partial f}{\partial y} e_2$$

# Optimization 101

## Gradient ascent

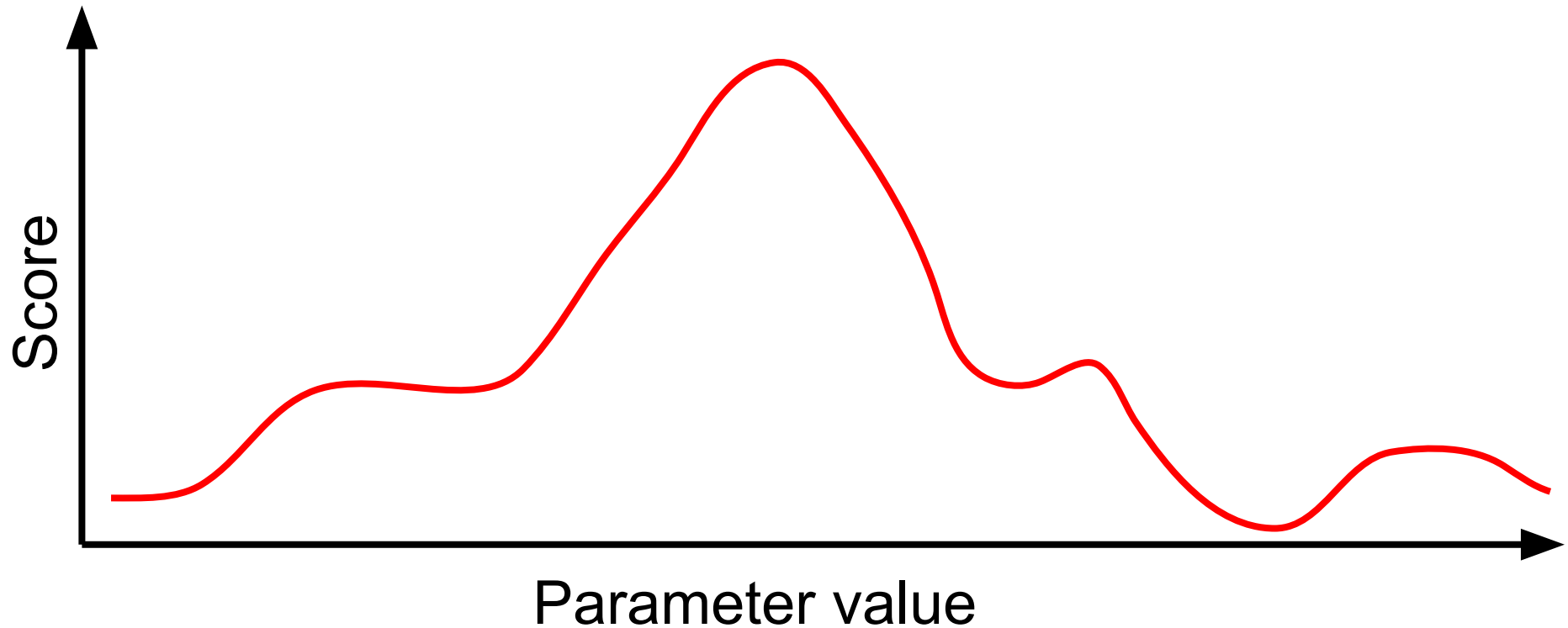


Search might fail if the search space isn't convex



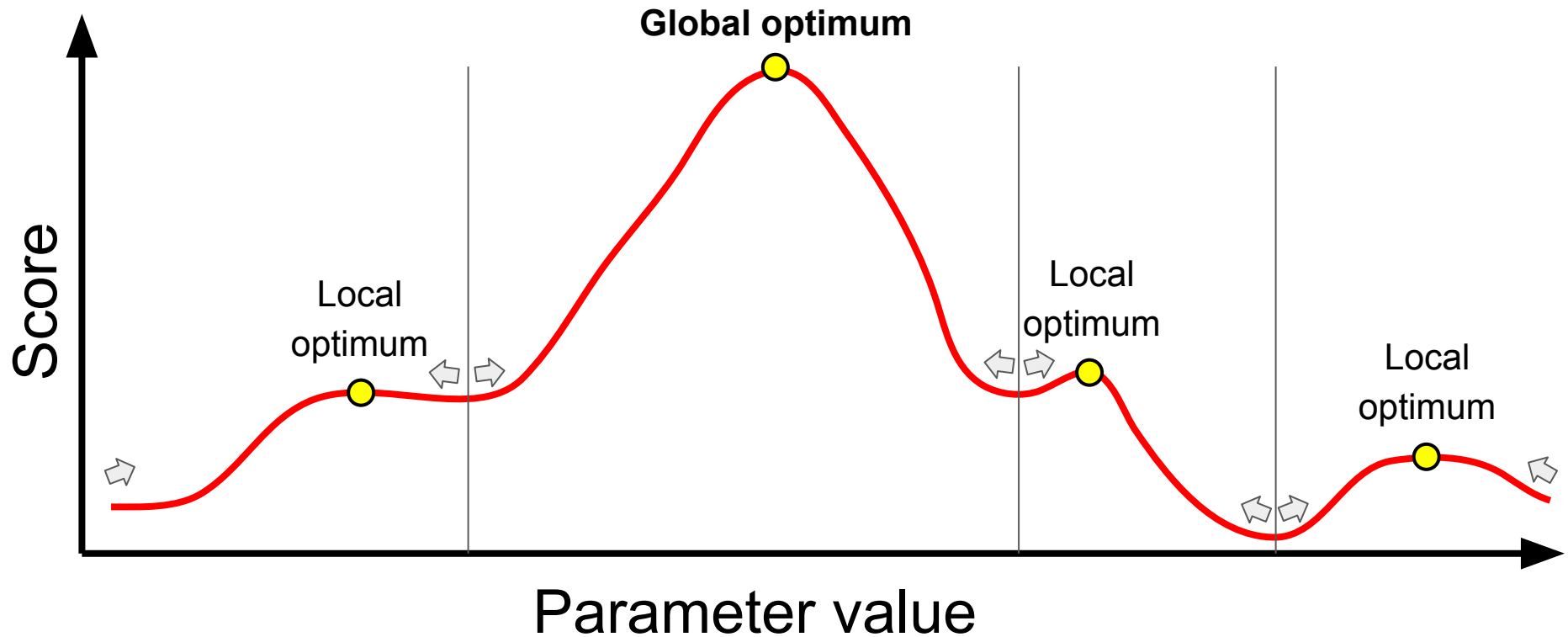
# Optimization 101

Search might fail if the search space isn't convex



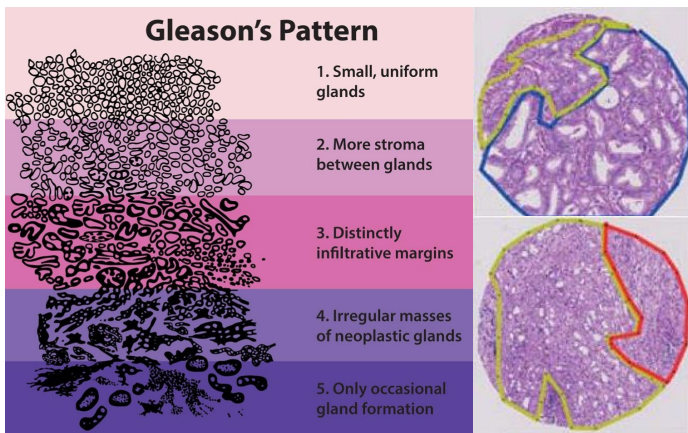
# Optimization 101

Search might fail if the search space isn't convex

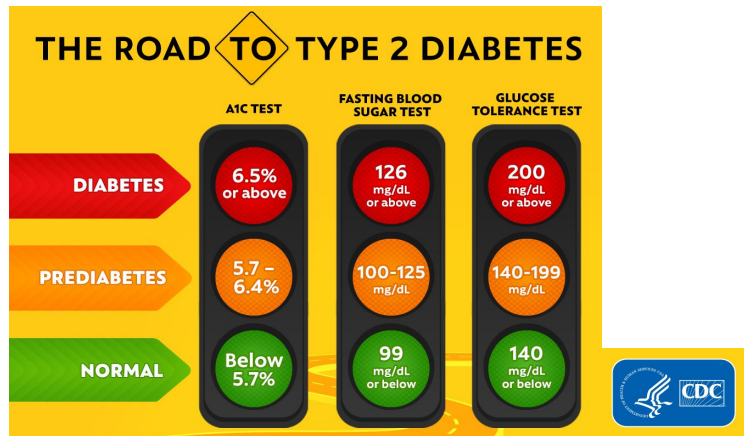


# Back to medicine - why should we care?

## Gleason grading system



## Prediabetes diagnosis



## Blood Oxygen levels

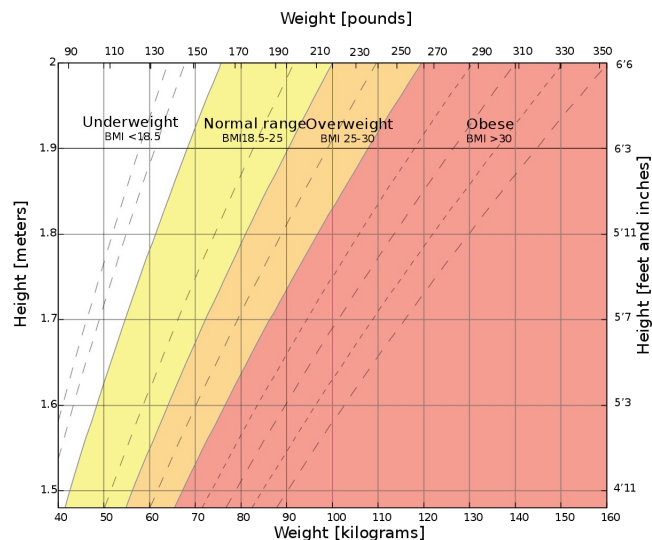


**Normal Blood Oxygen Levels Pulse Oximeter Chart**

95-100%	Normal Blood Oxygen Levels
91-95%	'Concerning' Blood Oxygen Levels
≥ 90%	Low Blood Oxygen Levels
80-85%	When Low Oxygen Saturation Affects Your Brain
87%	Cyanosis

© Carewell

## Body mass index



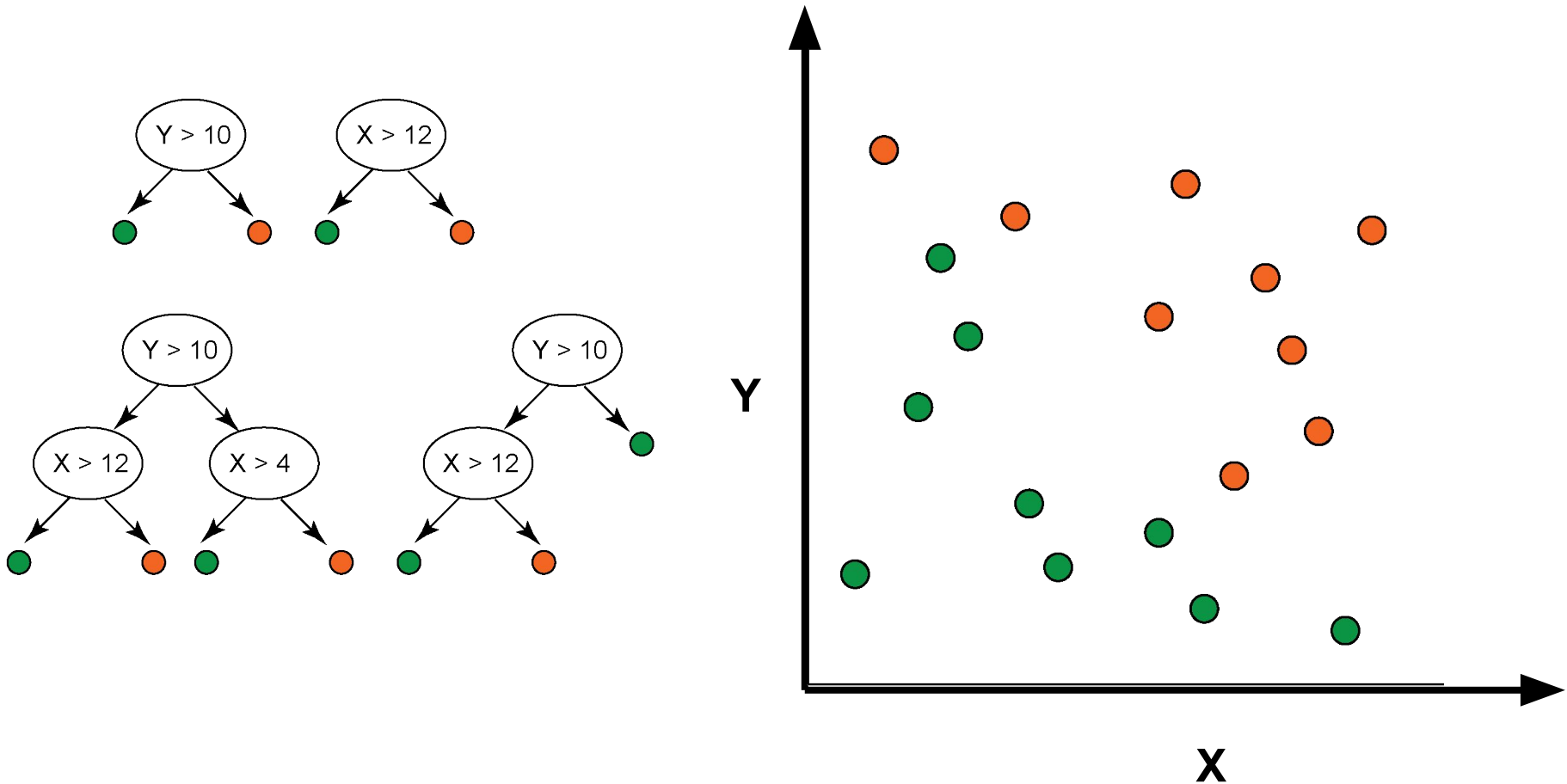
Or cholesterol, or breast cancer, or resting heart rate, etc.

# Advanced subjects

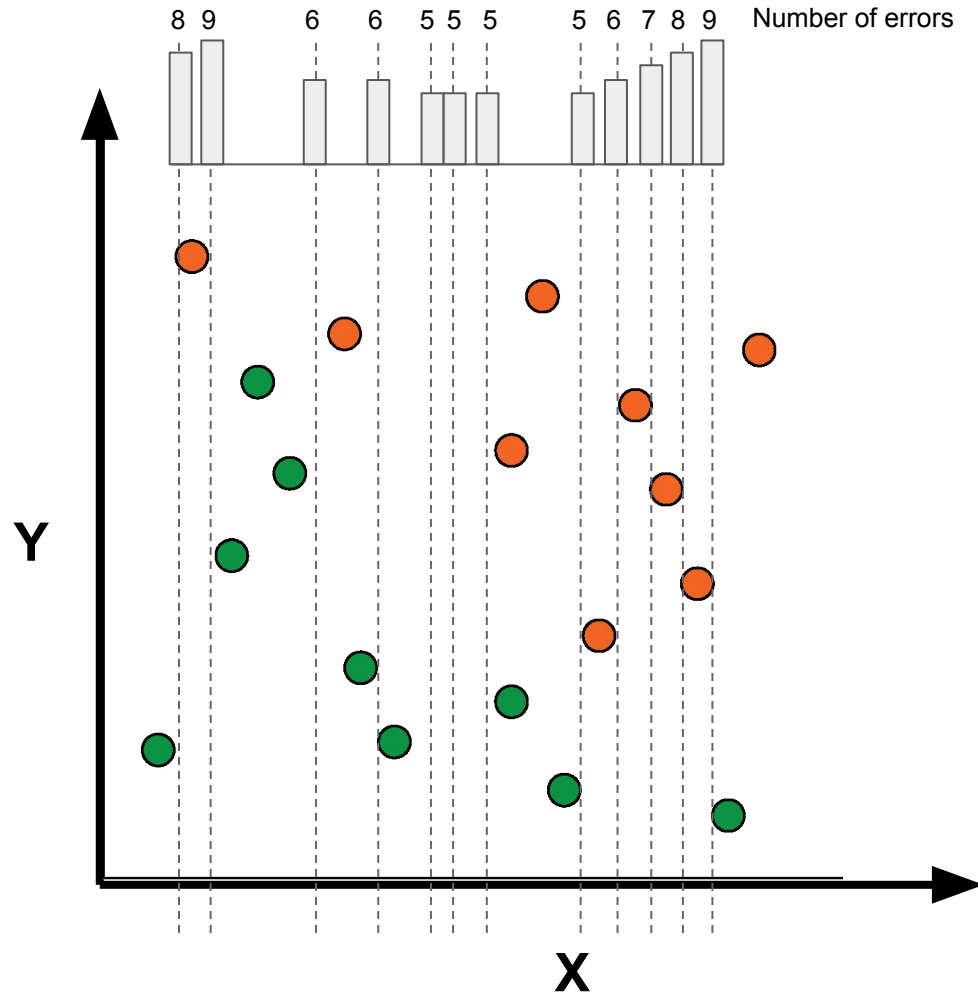
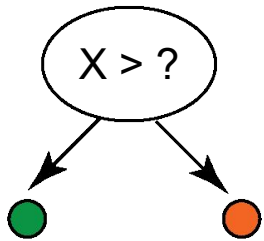
- Optimal solution vs Heuristics
- Convexity

# Discrete Search

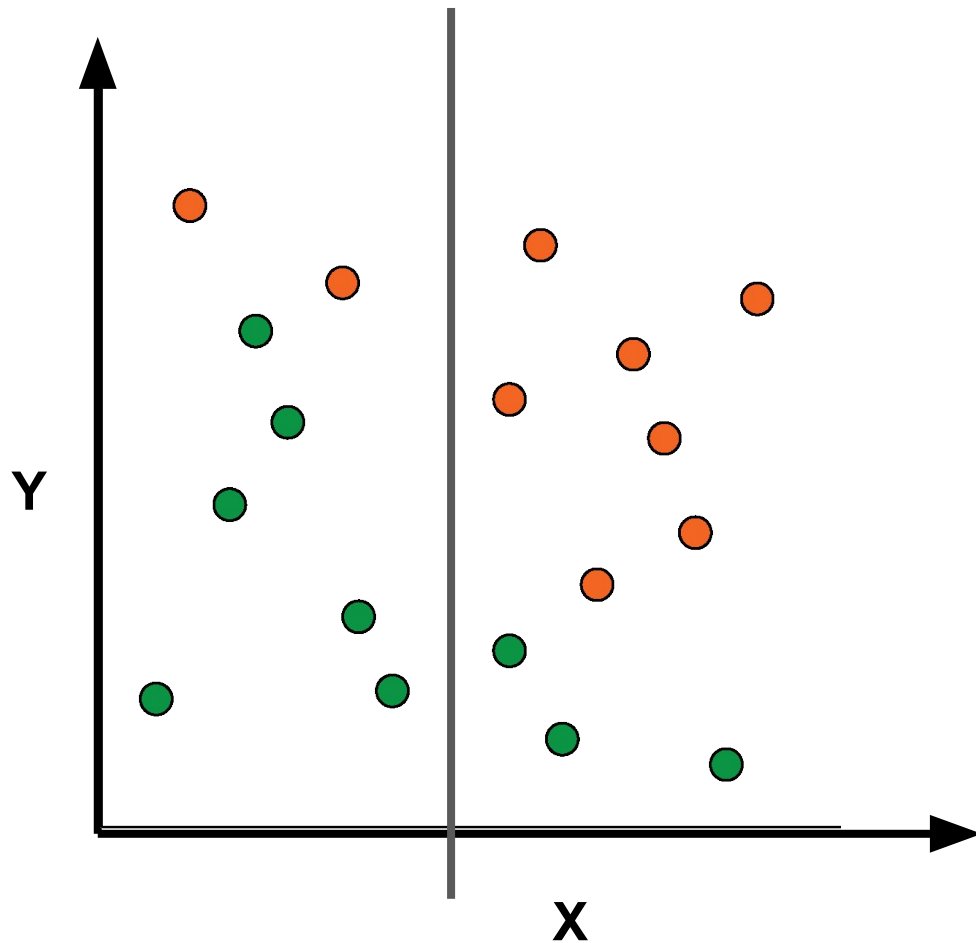
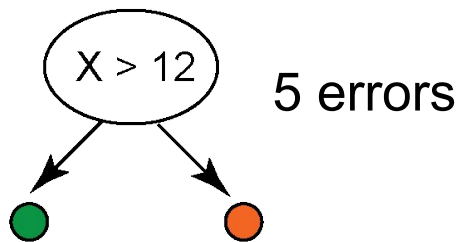
Learning decision trees?



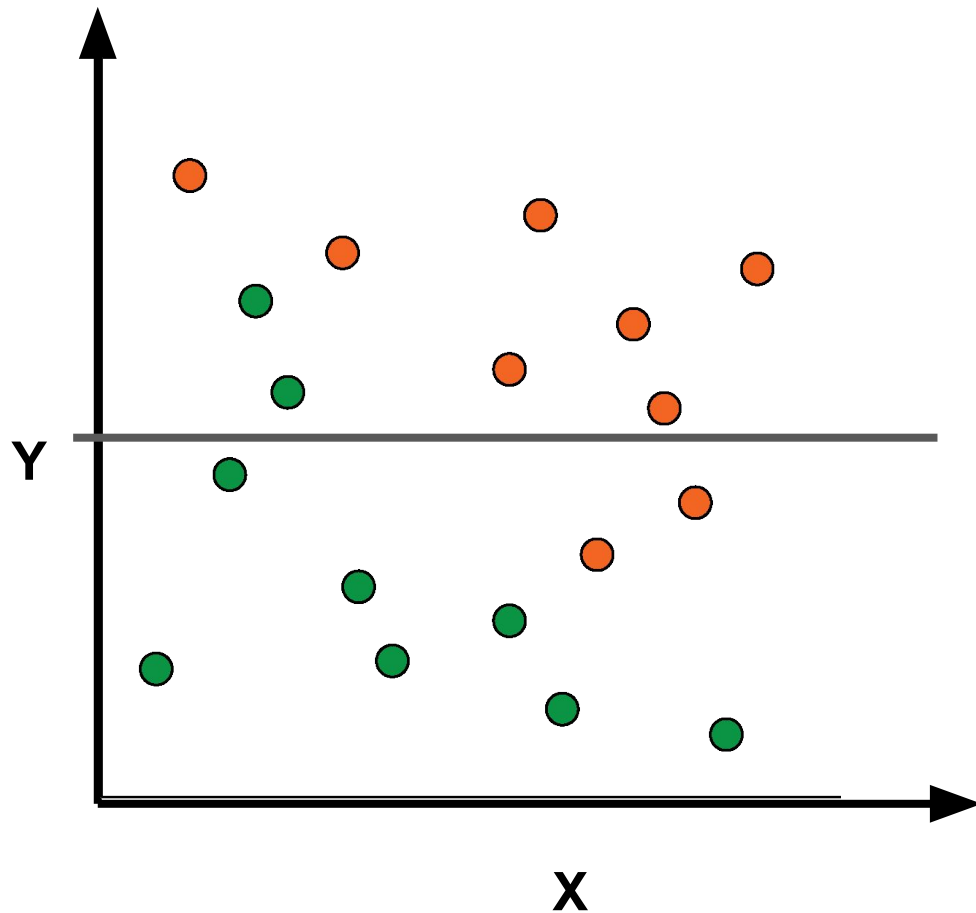
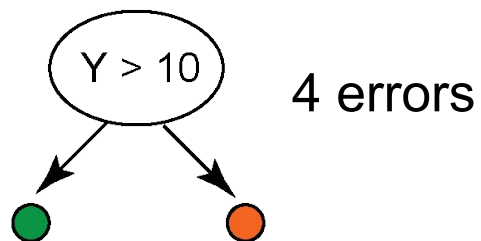
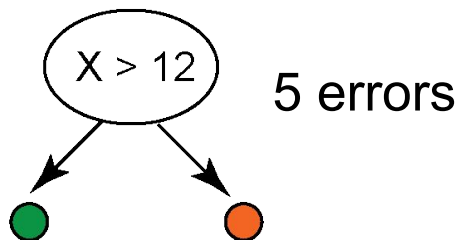
# Discrete Search



# Discrete Search

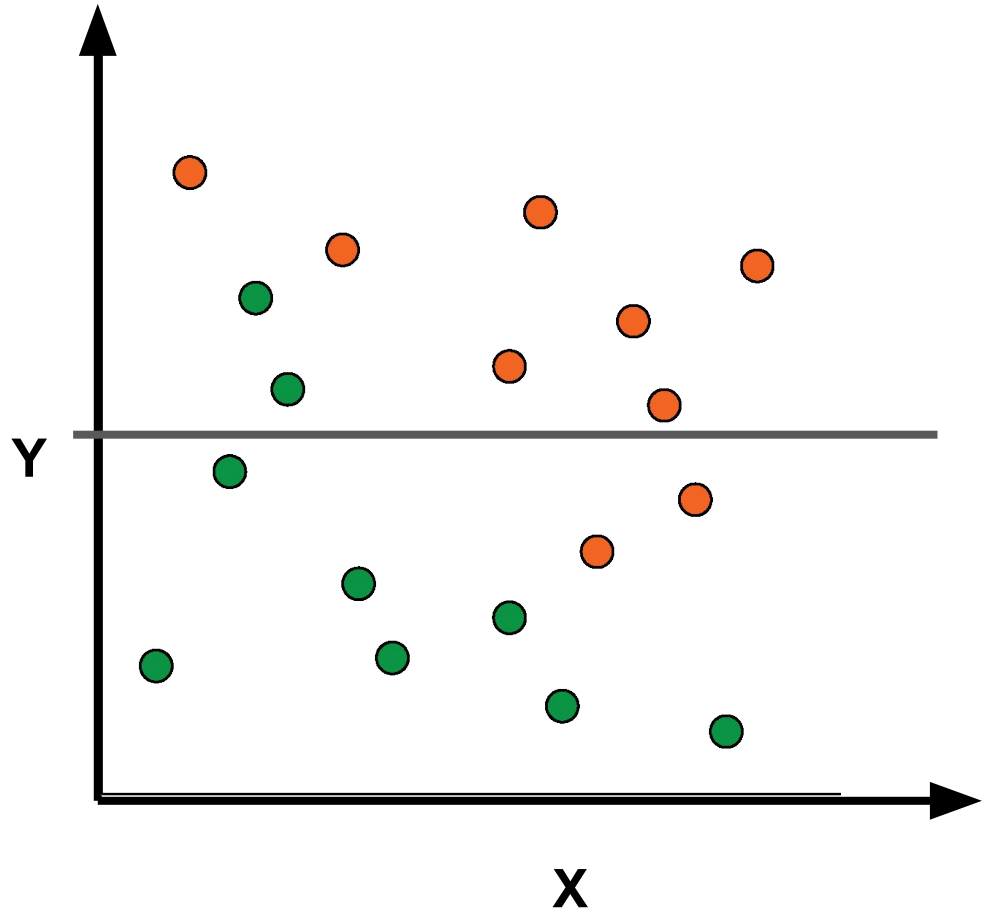
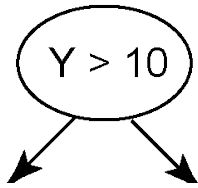


# Discrete Search

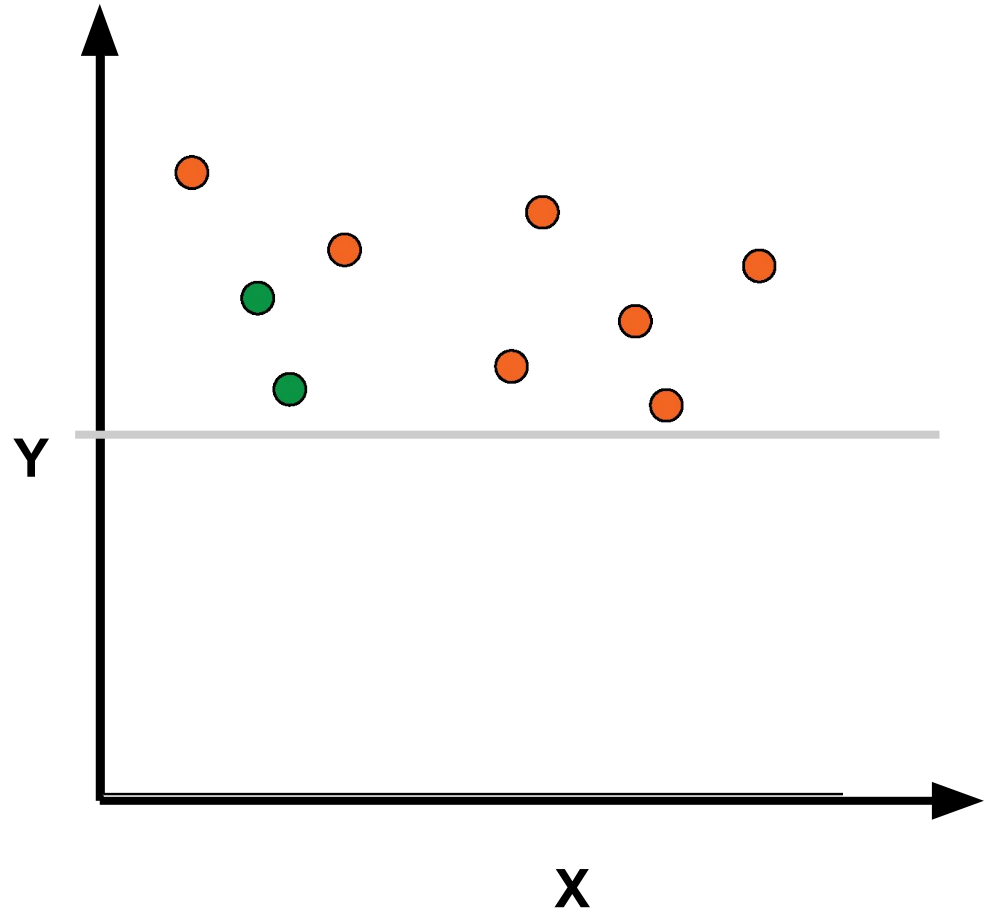
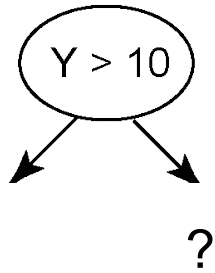




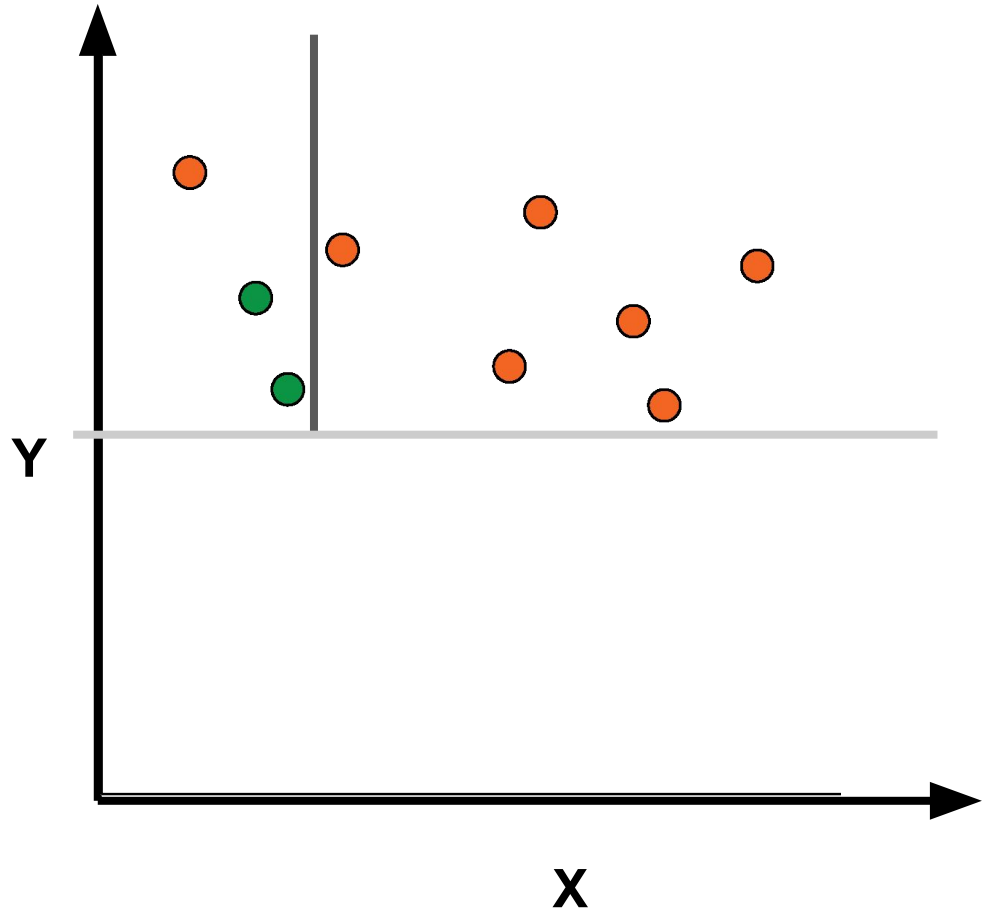
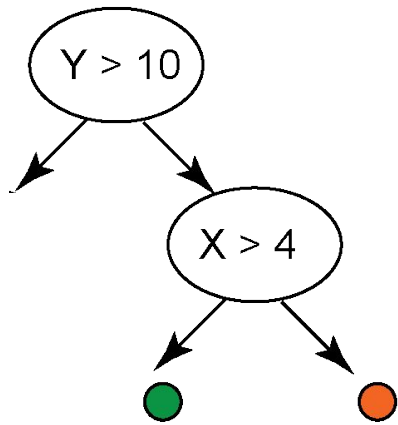
# Discrete Search



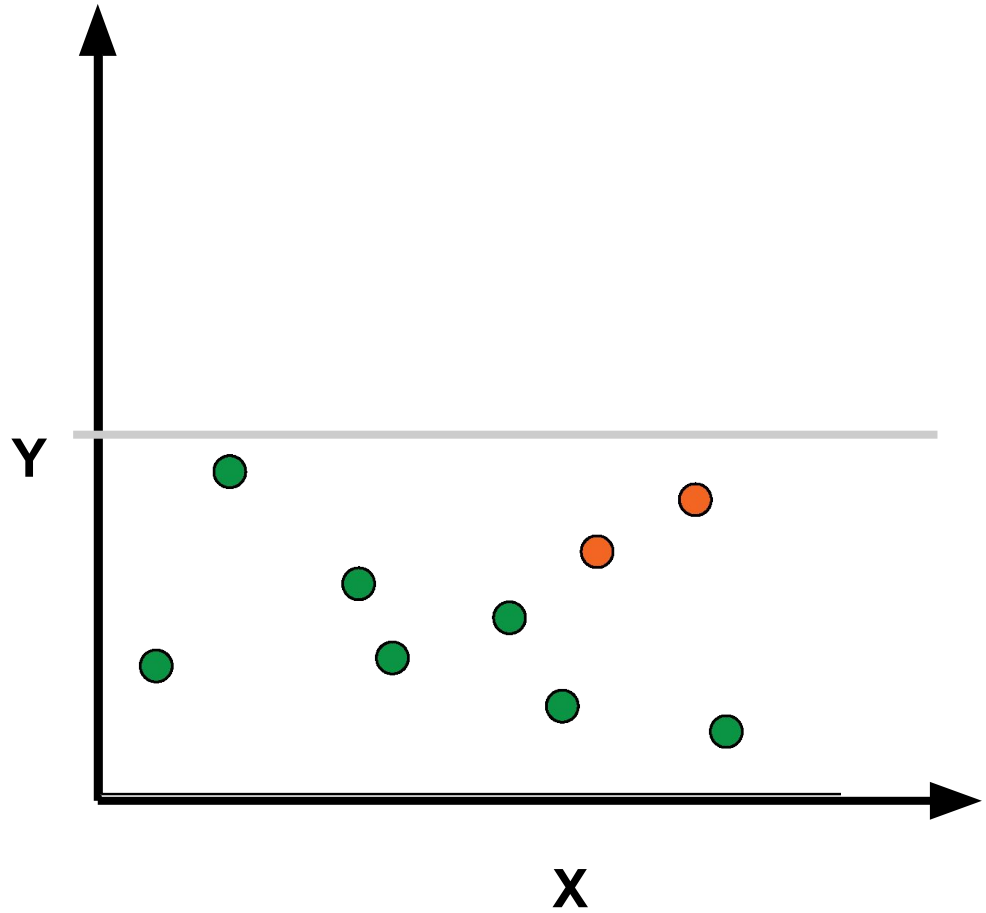
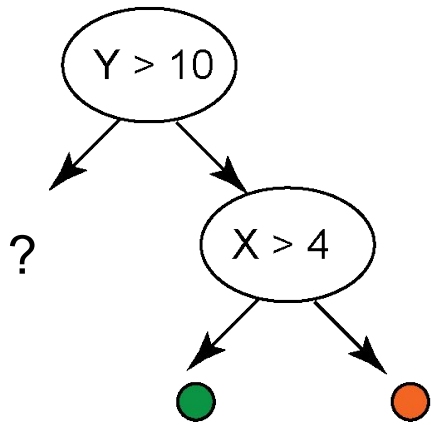
# Discrete Search



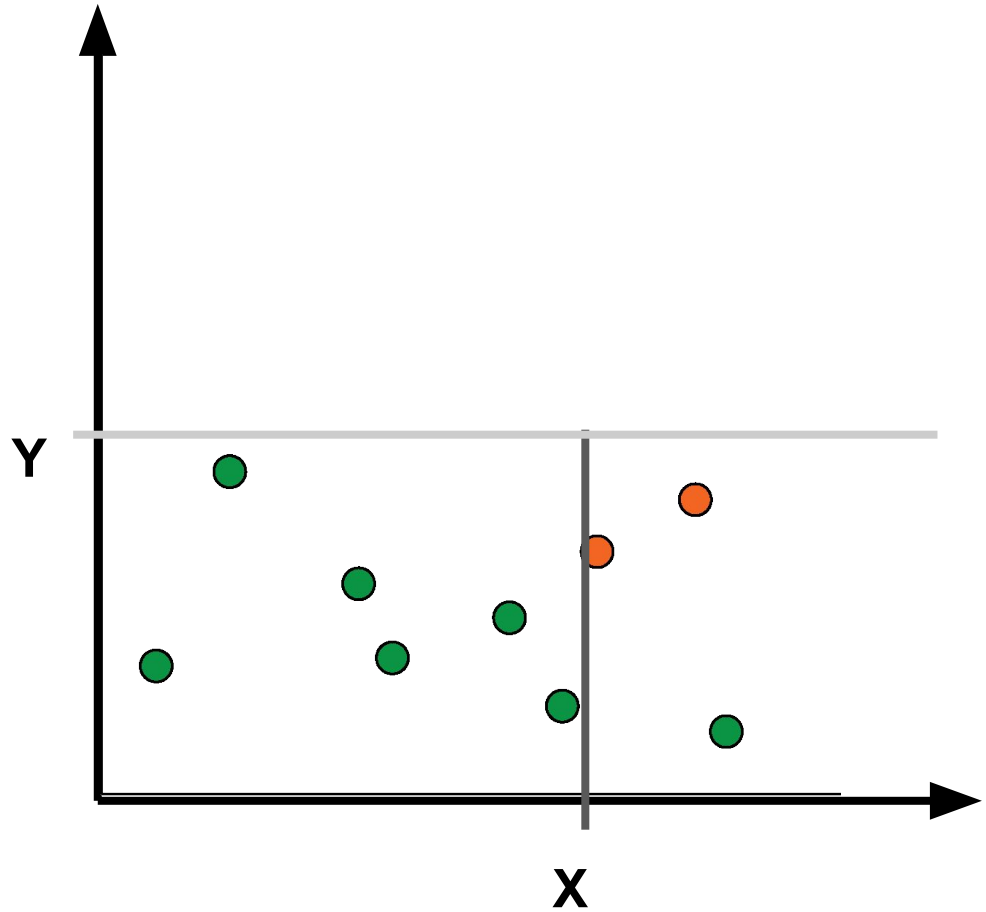
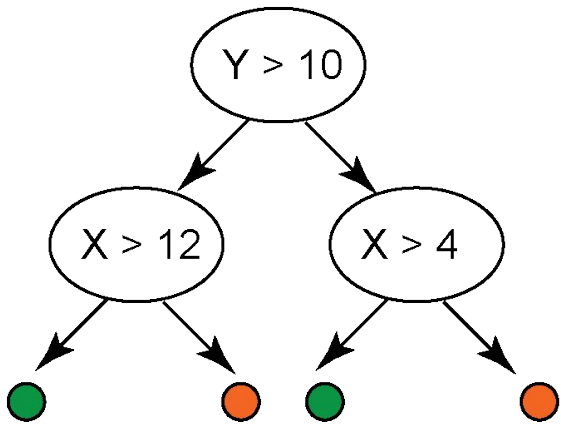
# Discrete Search



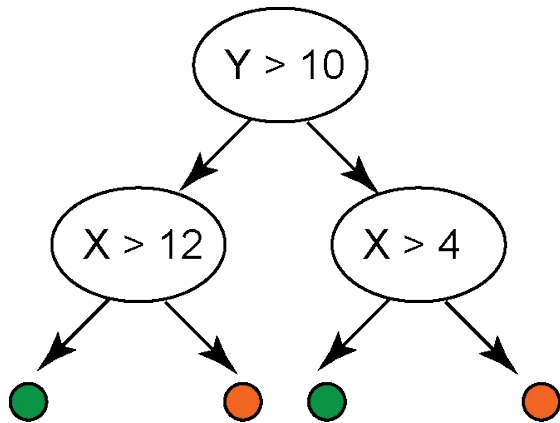
# Discrete Search



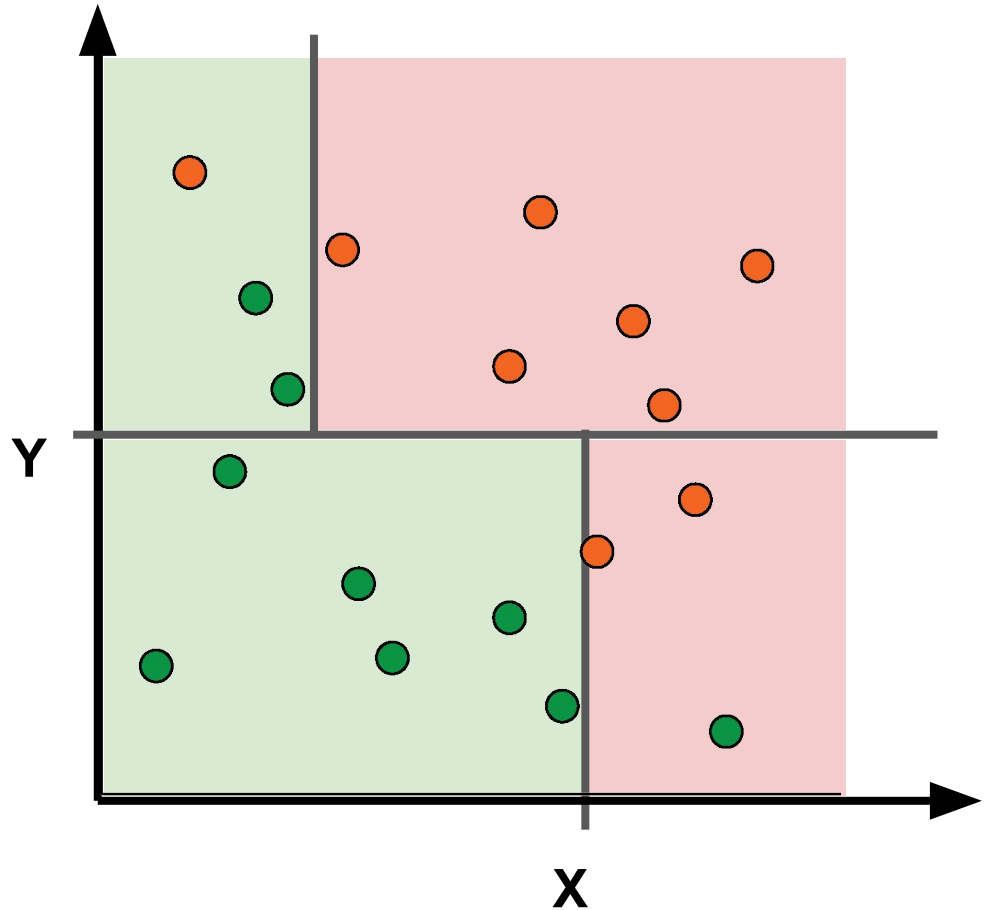
# Discrete Search



# Discrete Search



2 errors



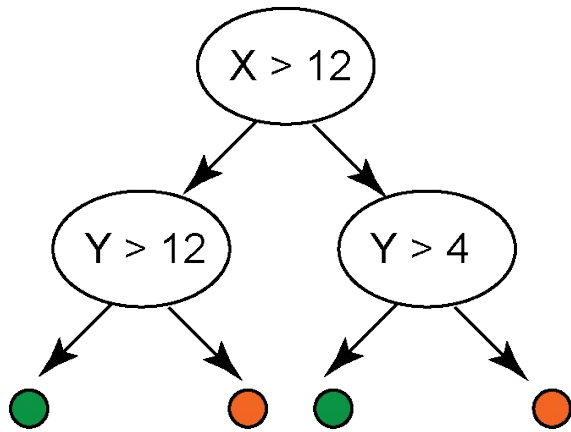
# Greedy Algorithms

"An algorithm which always takes the best immediate, or local, solution while finding an answer"

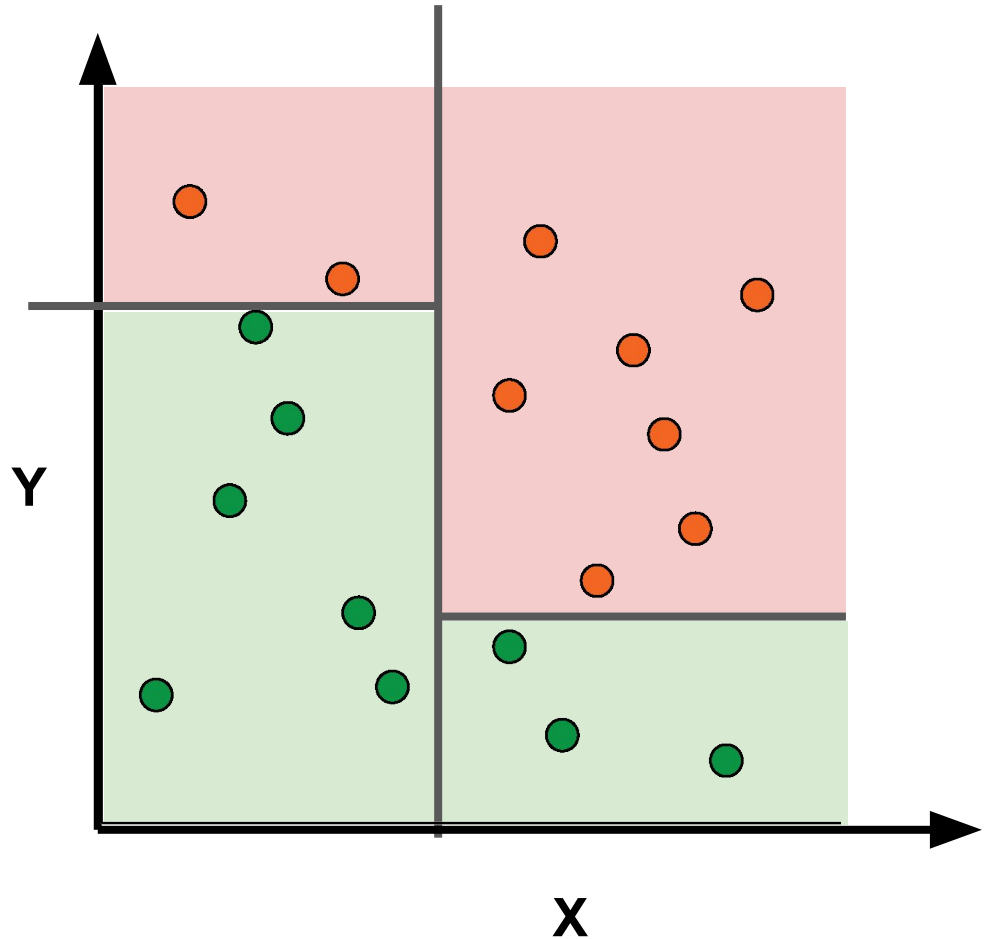
(from <http://www.nist.gov/dads/HTML/greedyalgo.html>)

- + Simple, local decisions
- May not reach global optimum

# Discrete Search



0 errors





# Summary

- Learning (rules) from data
- Parameter space and hypothesis space
- Loss function (how much errors cost)
- Finding the parameters of the optimal classifier
- Line search optimization
- Gradient descent (and convex optimization)
- Greedy algorithms and heuristics

The basics of machine learning