Bradley Voytek, Ph.D.

UC San Diego

Department of Cognitive Science
Halıcıoğlu Data Science Institute
Neurosciences Graduate Program

bvoytek@ucsd.edu
@bradleyvoytek

COGS 108
Data Science in Practice

ML basics

UC San Diego

What is machine learning?

"Machine Learning (ML) is a fascinating field of artificial intelligence (AI) research and practice where we investigate how computer agents can improve their perception, cognition, and action with experience. Machine Learning is about machines improving from data, knowledge, experience, and interaction." - Manuela Veloso, Head of ML at Carnegie Mellon

What is machine learning?

"Machine learning is the science of getting computers to act without being explicitly programmed" - Andrew Ng, Stanford, ex-Google, formerly chief scientist at Baidu, founder Coursera

What is machine learning?

"Machine learning is a method of data analysis that automates analytical model building." - SAS, the largest installed user base of analytics software, revenue of $3.2 billion/year

What does ML look like?

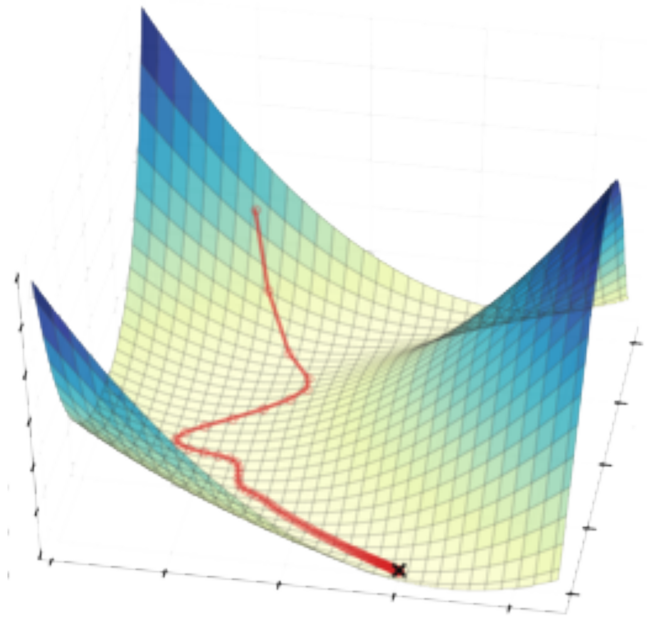
- Algorithms that can improve in performance with training or experience
- Typically by fitting large numbers of parameters
- Usually used in situations where it is difficult to define rules by hand
  - Face detection
  - Stock market prediction
  - Spam filters

What is ML really?

What is ML really?

What is ML really?

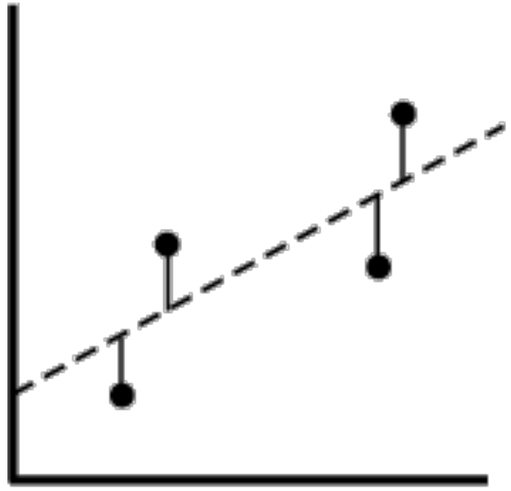
Pick a set of parameter values that minimize some objective function



What is ML really?

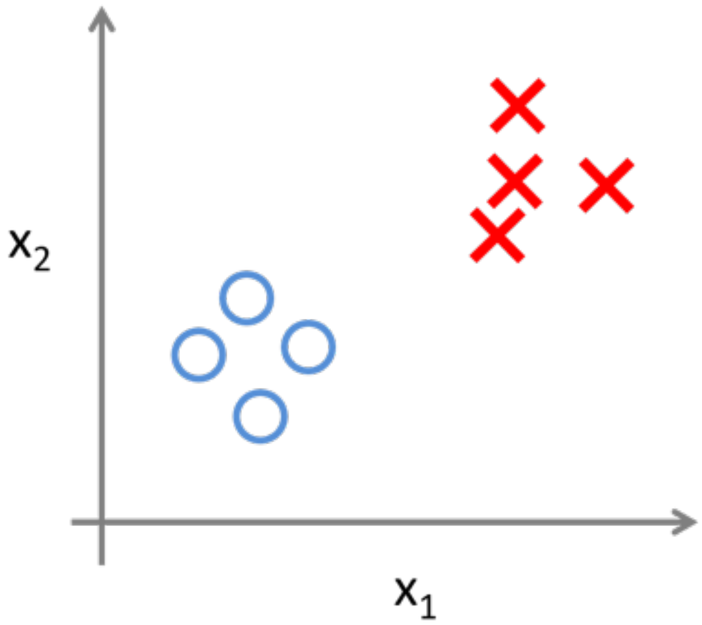
Pick a set of parameter values that minimize some objective function

$$\text{argmin}_{a,b} \sum_{i=1:n} (y_i - (ax_i + b))^2$$

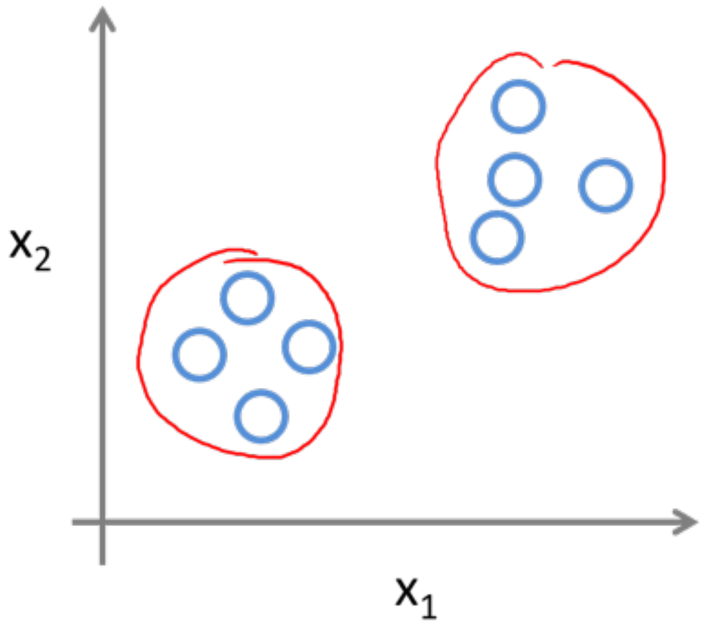


Two major modes of ML

Supervised Learning

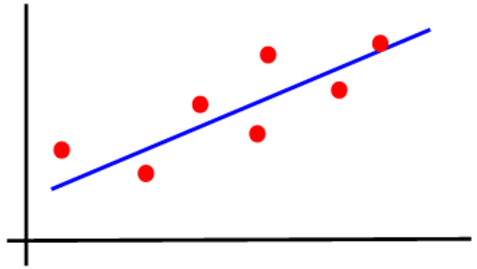


Unsupervised Learning



Three canonical problems

1. Regression - supervised
   - estimate parameters, e.g. of weight vs height

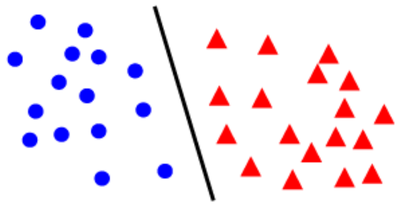


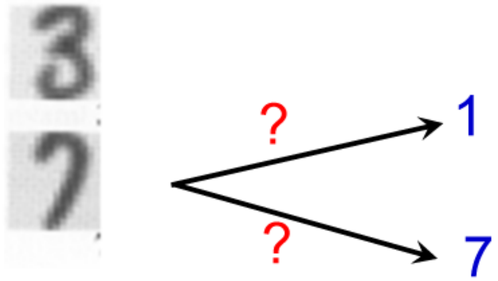
Source: Andrew Zisserman

Three canonical problems

1. Regression - supervised
   - estimate parameters, e.g. of weight vs height

2. Classification - supervised
   - estimate class, e.g. handwritten digit classification





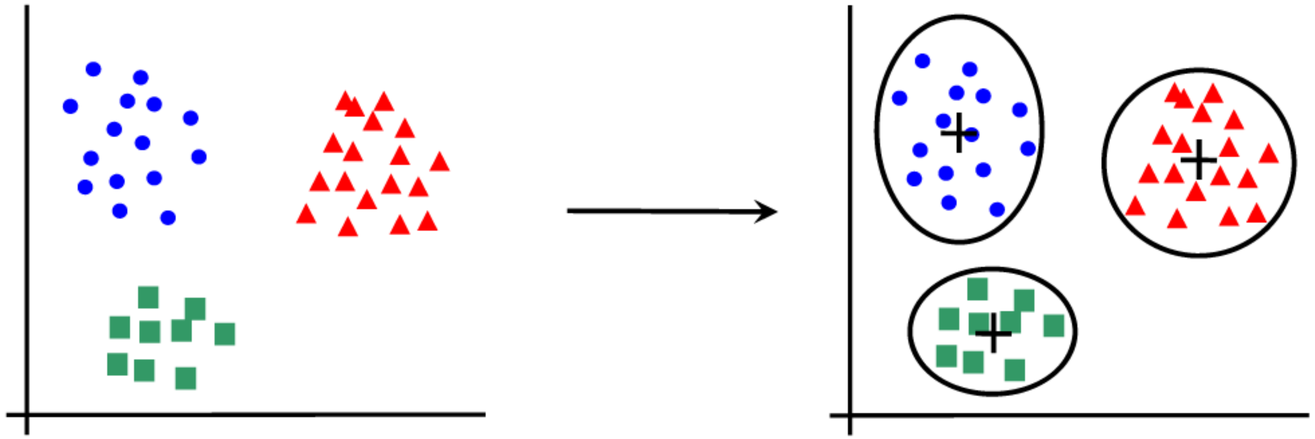
Source: Andrew Zisserman

Three canonical problems

1. Regression - supervised
   - estimate parameters, e.g. of weight vs height

2. Classification - supervised
   - estimate class, e.g. handwritten digit classification

3. Unsupervised learning – model the data
   - clustering



Source: Andrew Zisserman

UC San Diego

COGS 108
Data Science in Practice

Support Vector Machines

UC San Diego

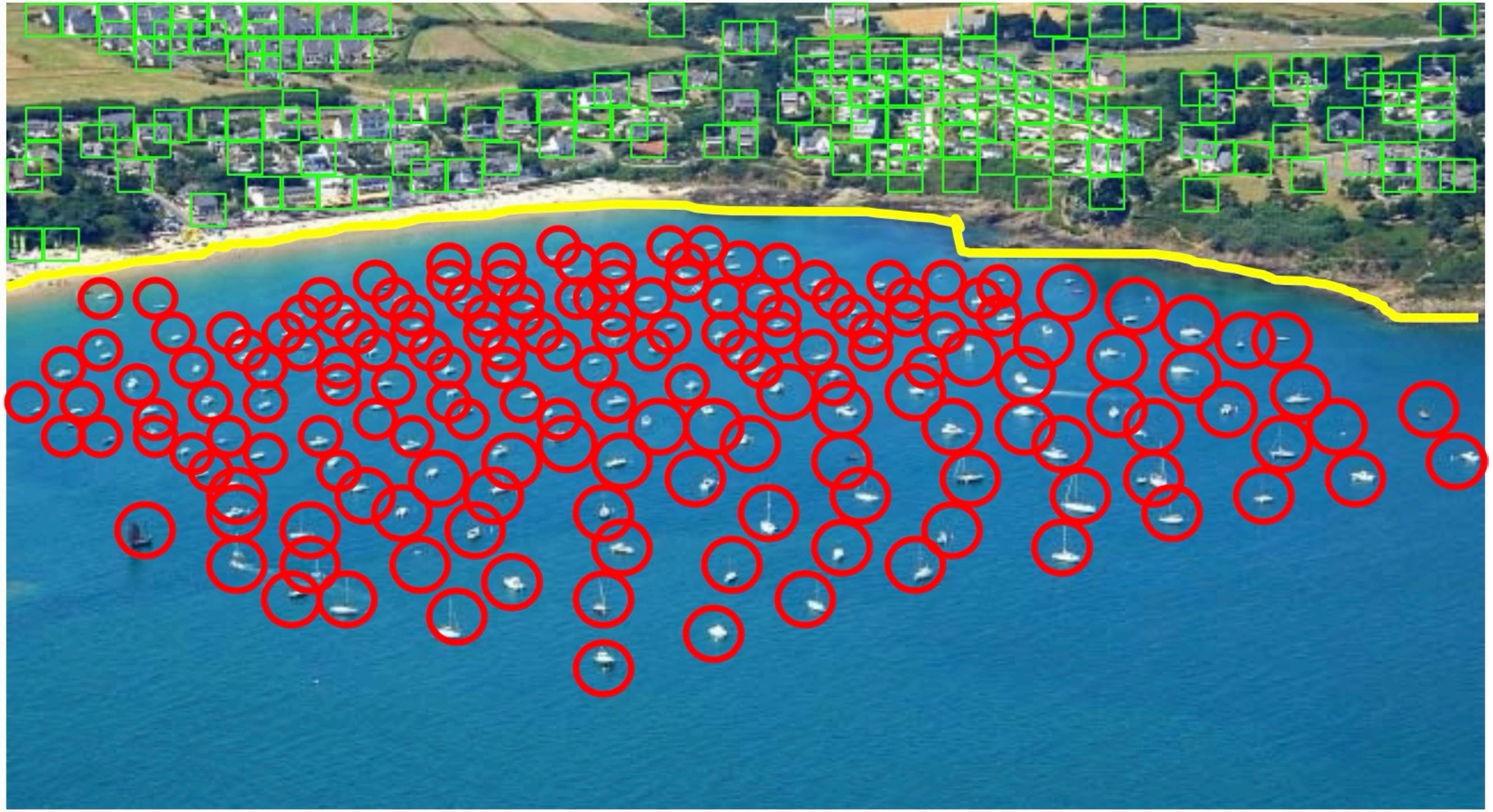
SVM

When classifying isn't easily linear,
***make it linear*.**

SVM

Source: Statnikov, Hardin, Guyon, Aliferis *AMIA* 2009

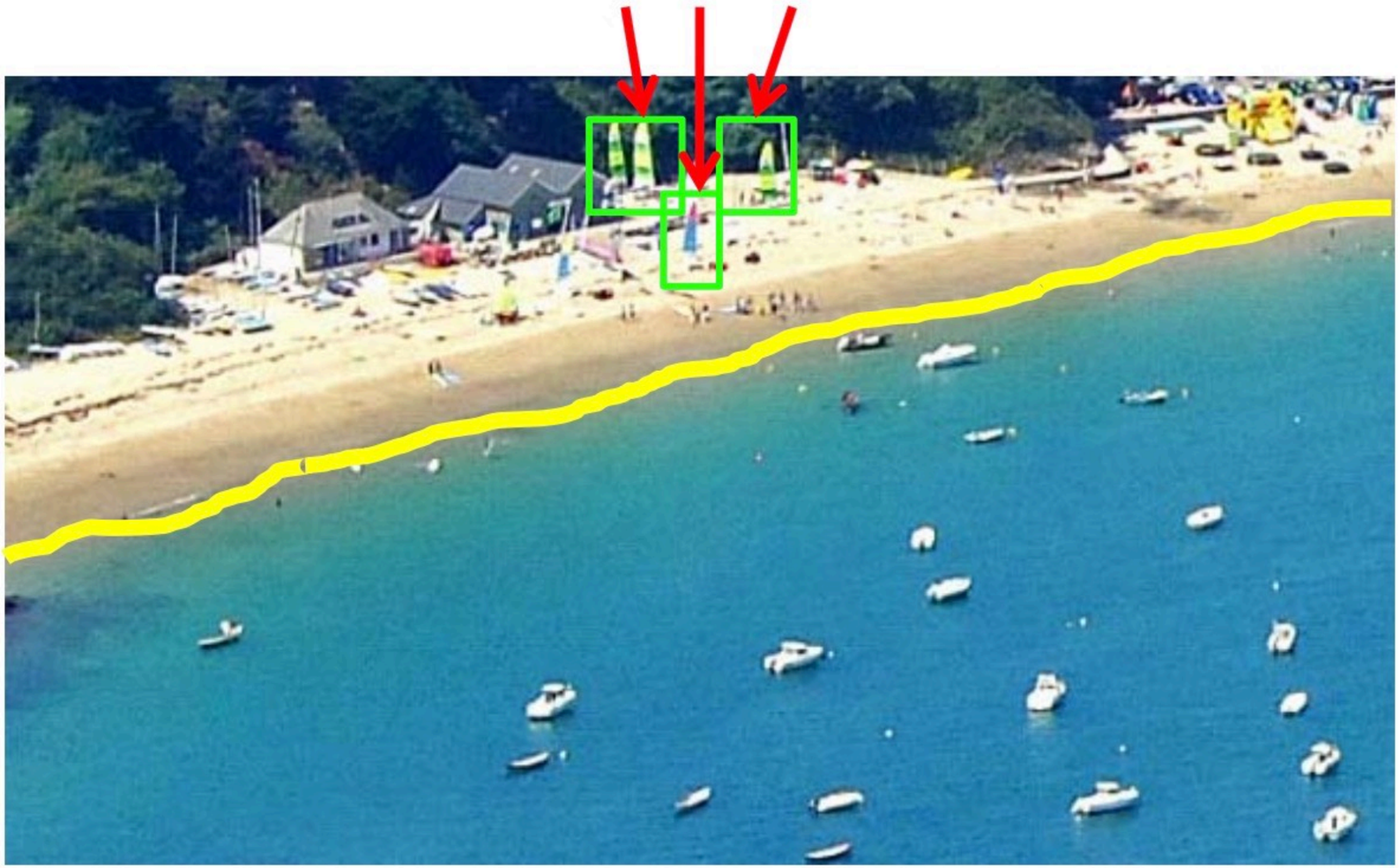
SVM



Source: Statnikov, Hardin, Guyon, Aliferis *AMIA* 2009

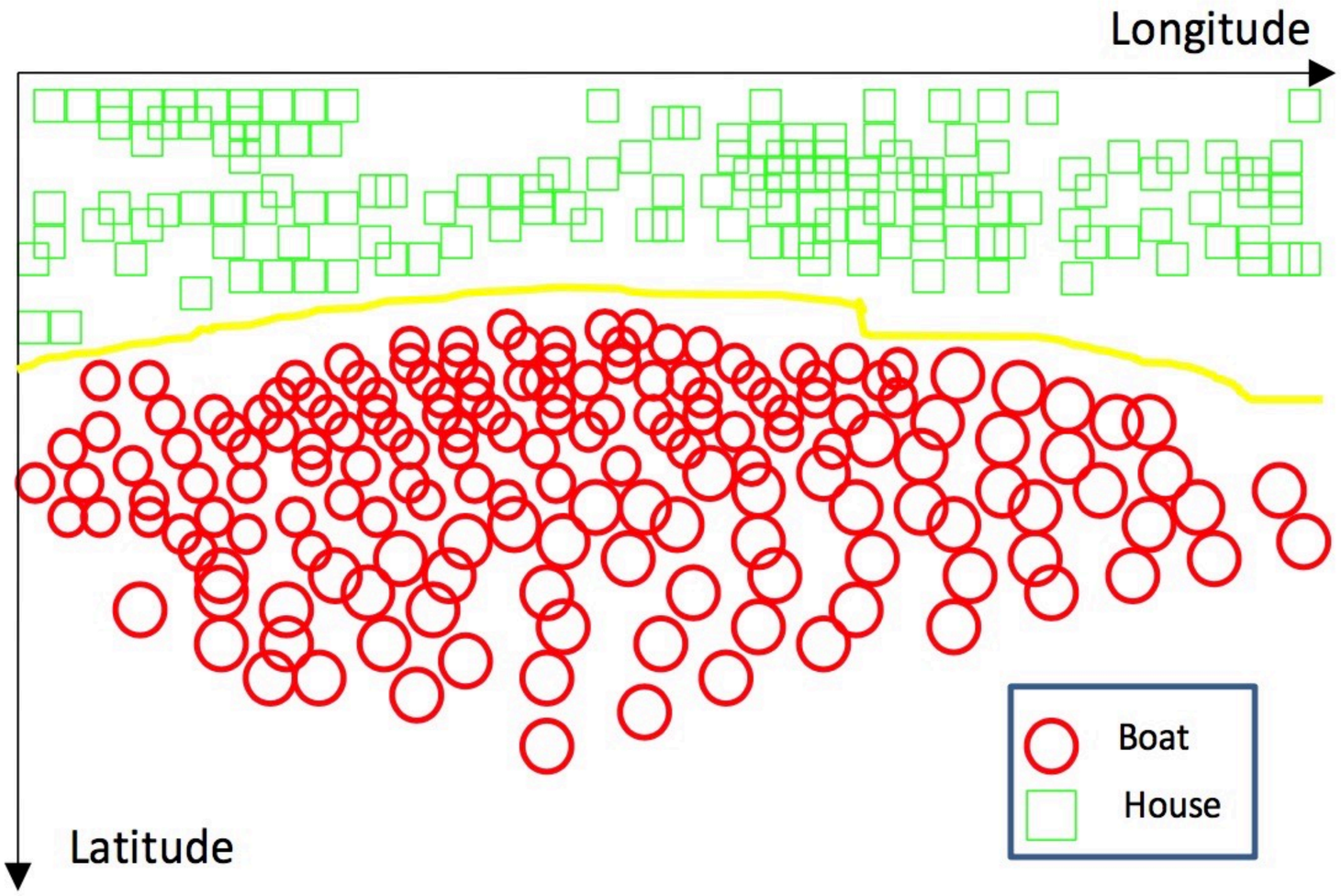
SVM

These boats will be misclassified as houses



Source: Statnikov, Hardin, Guyon, Aliferis *AMIA* 2009

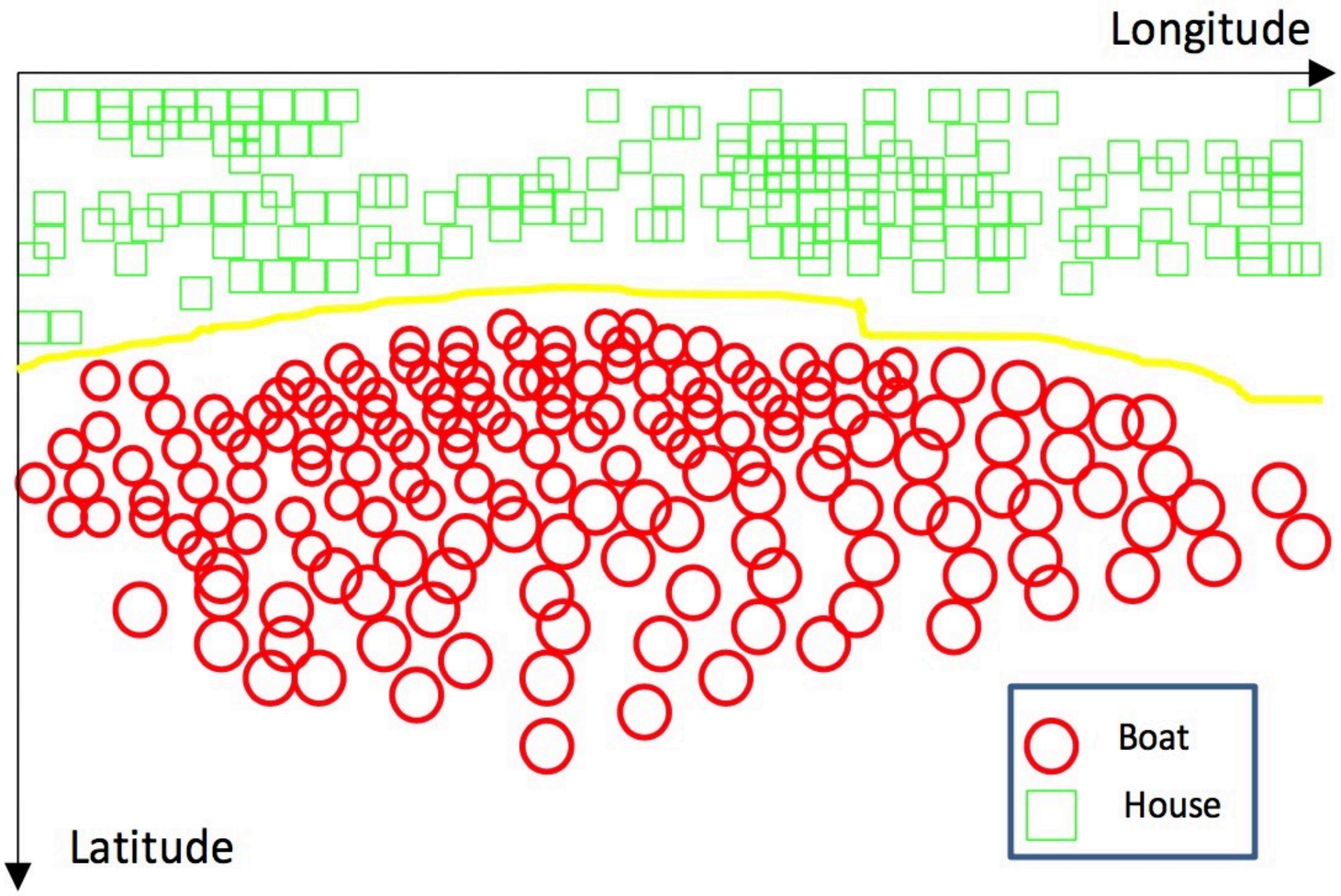
SVM



Shoreline is your decision boundary

Source: Statnikov, Hardin, Guyon, Aliferis *AMIA* 2009

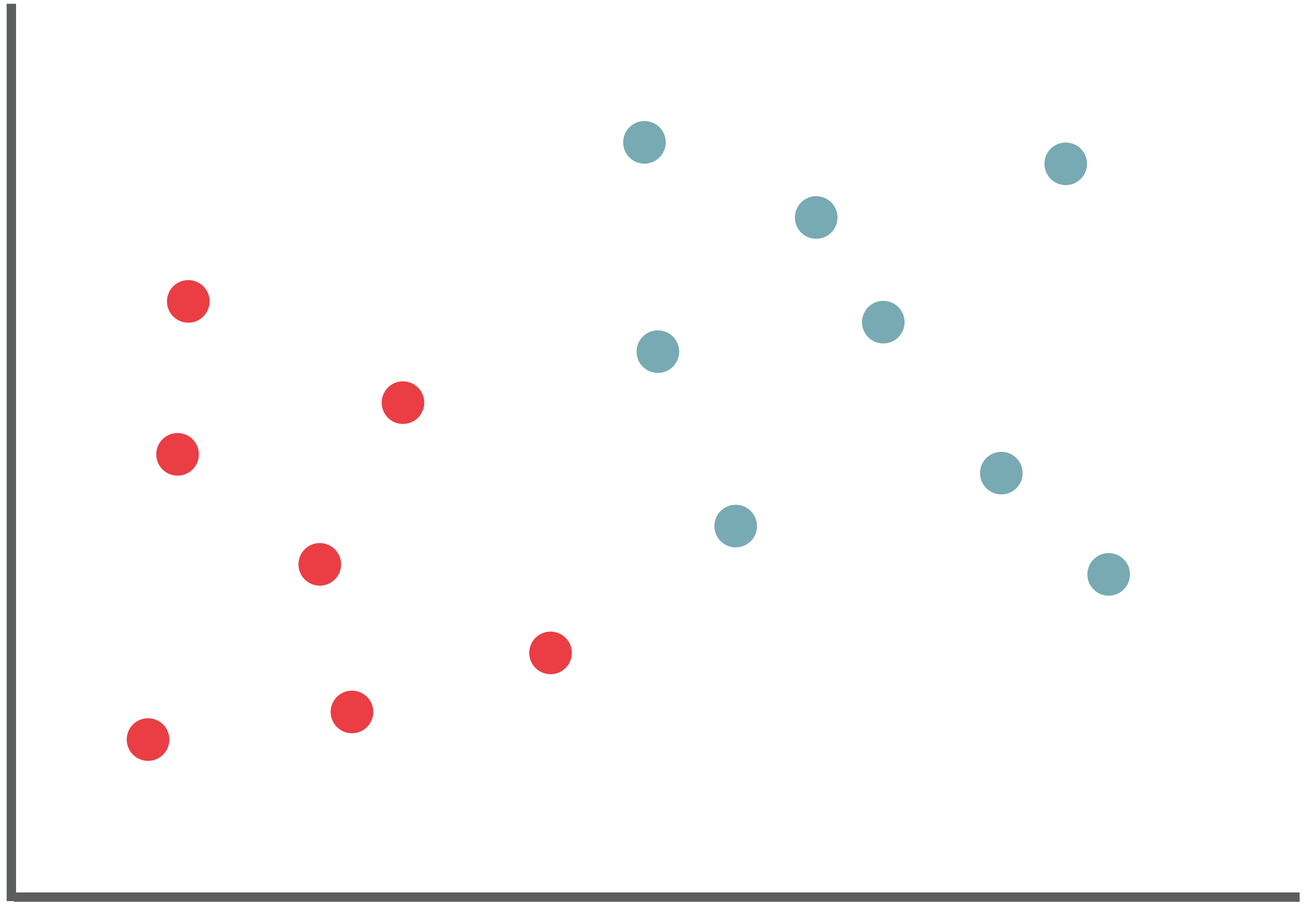
SVM



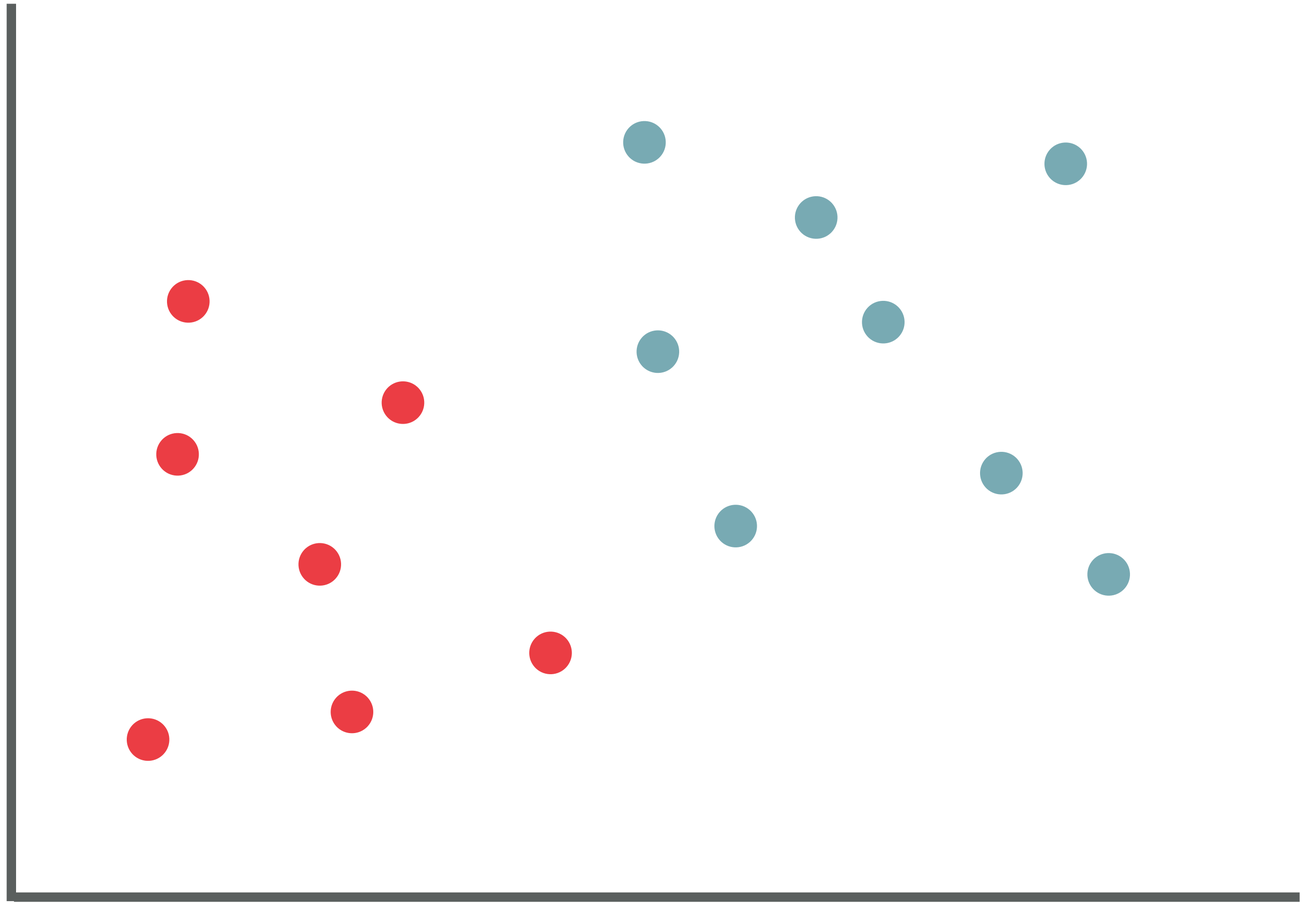
NOT LINEAR

Source: Statnikov, Hardin, Guyon, Aliferis *AMIA* 2009

SVM

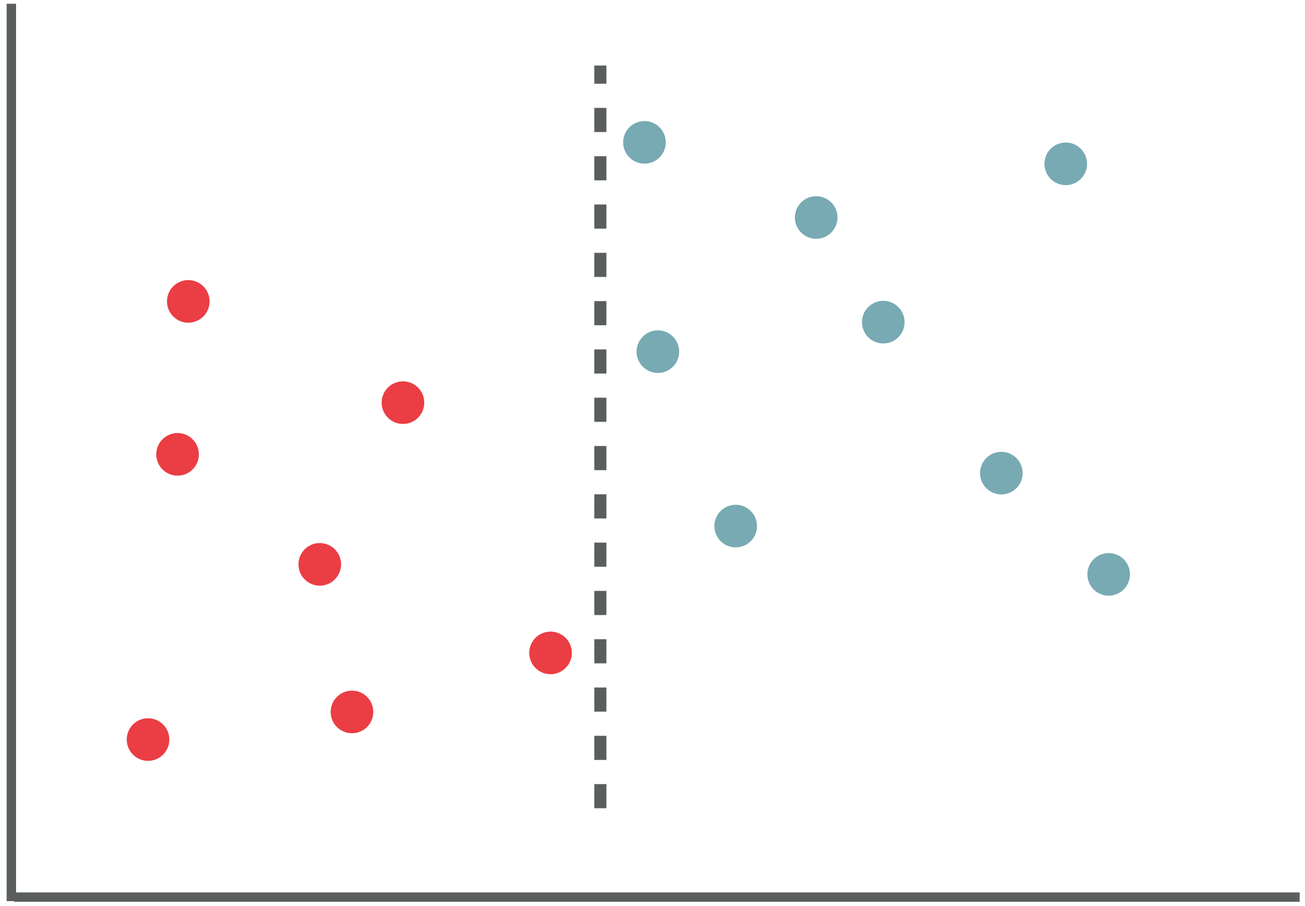


SVM



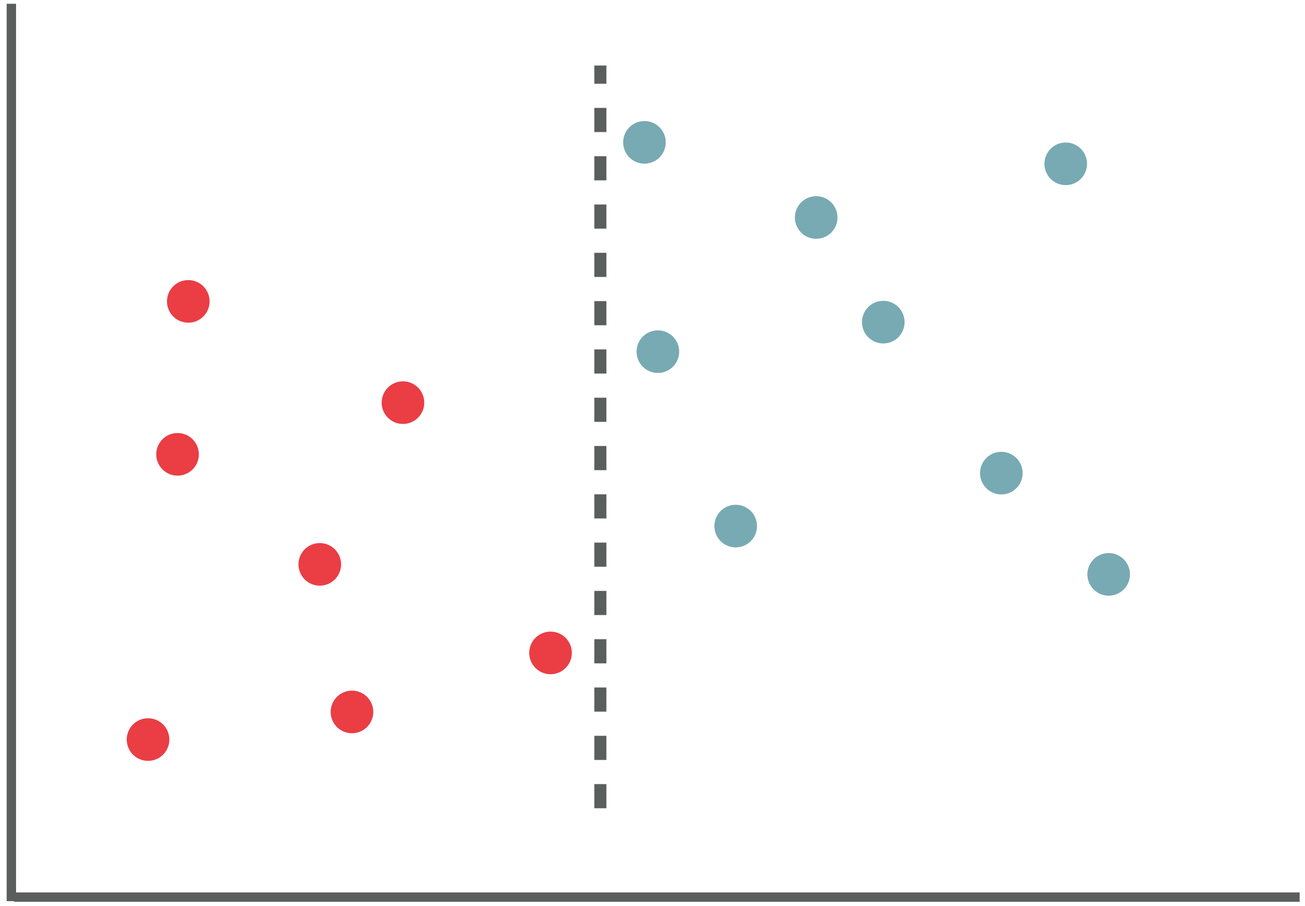
Classify!

SVM



Classify!

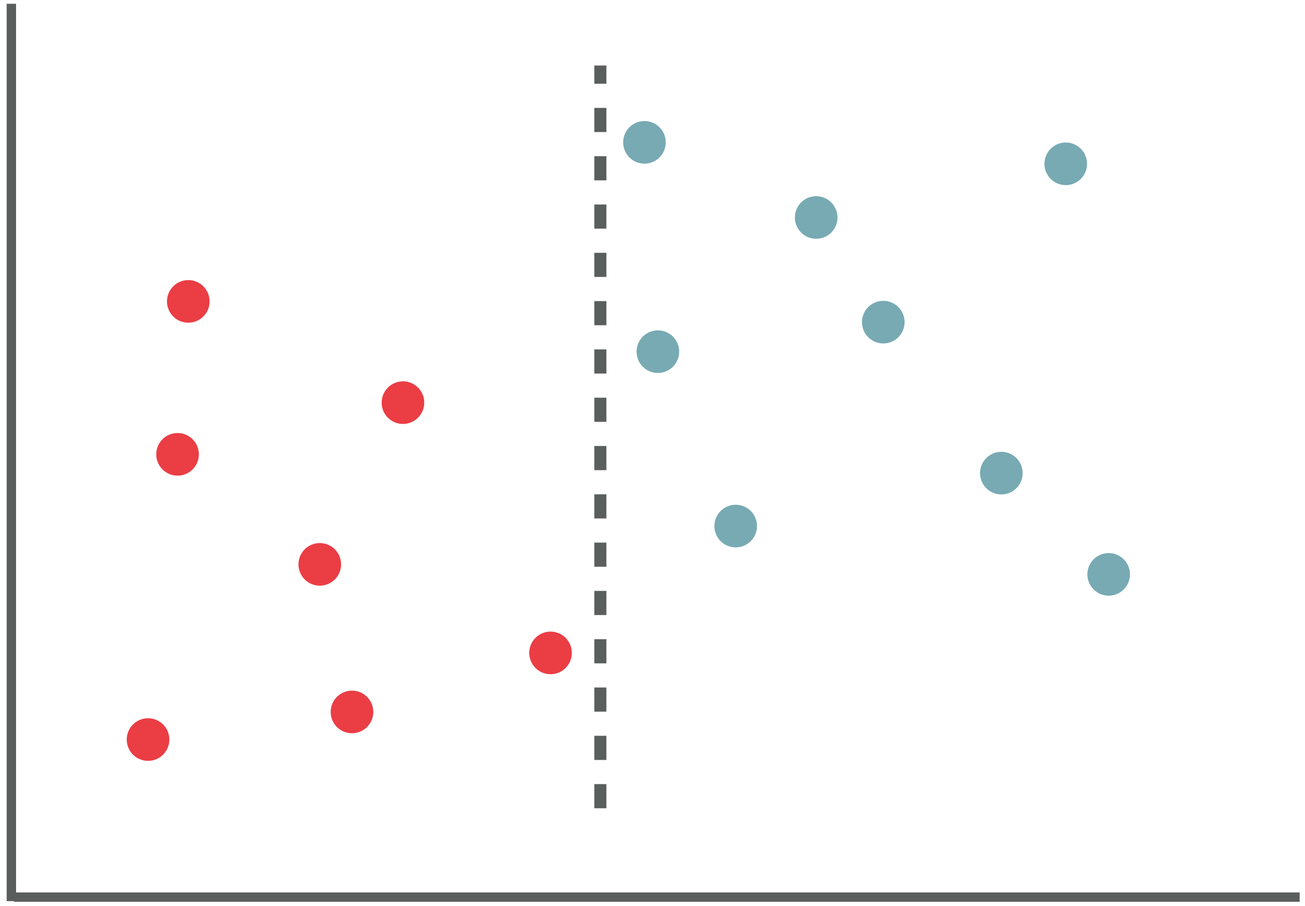
SVM



Classify!

DONE

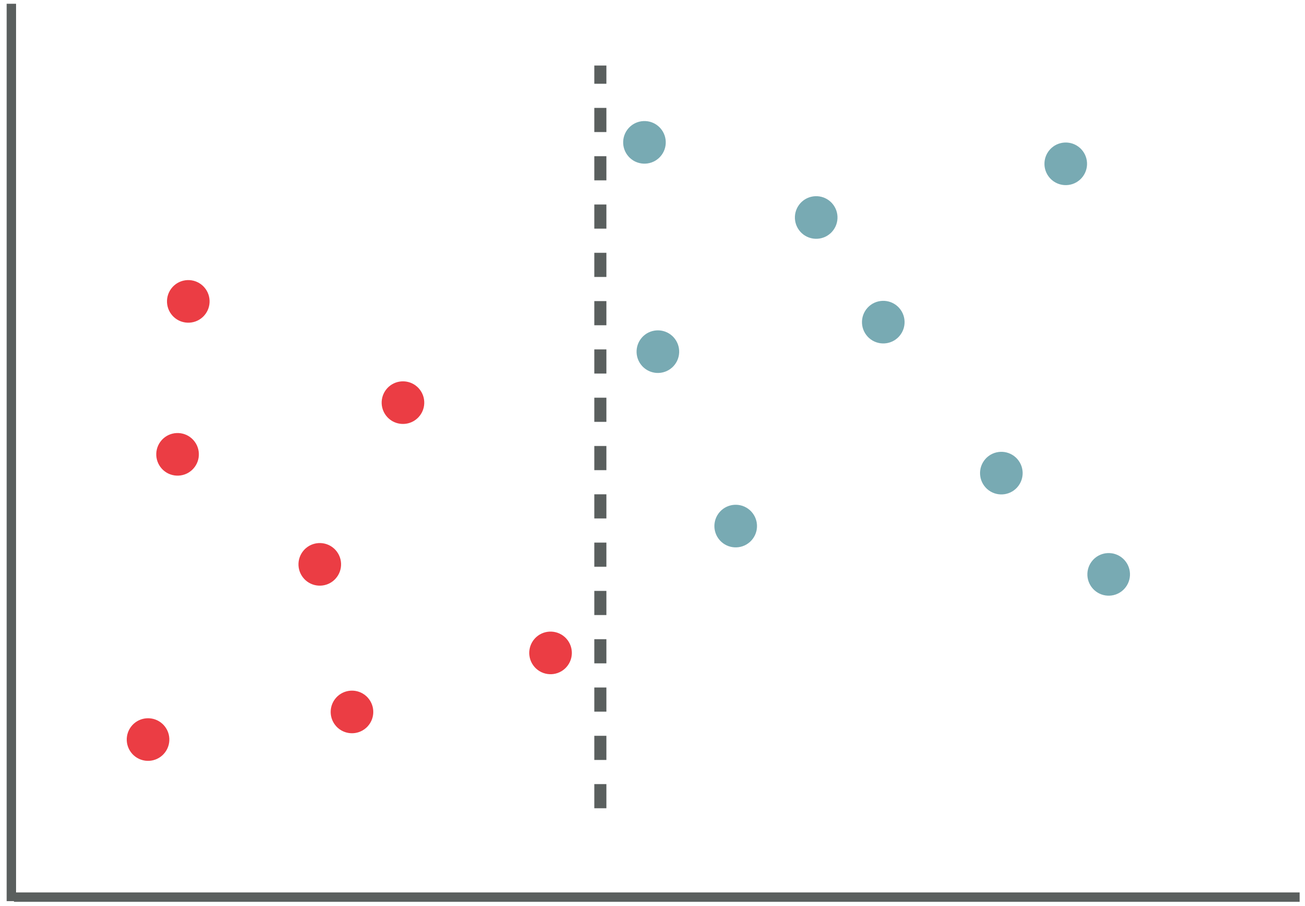
SVM



Classify!

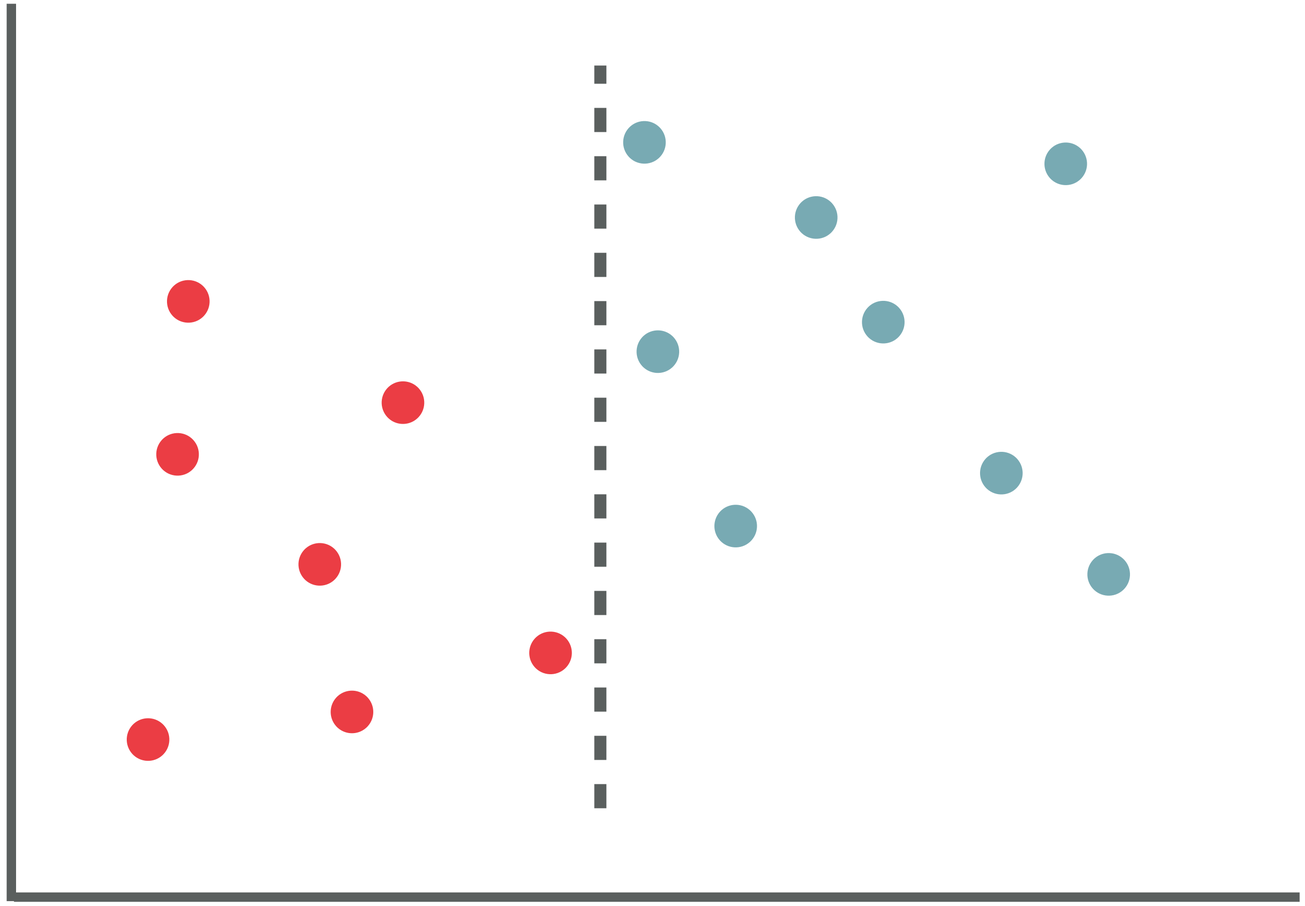
~~DONE~~

SVM



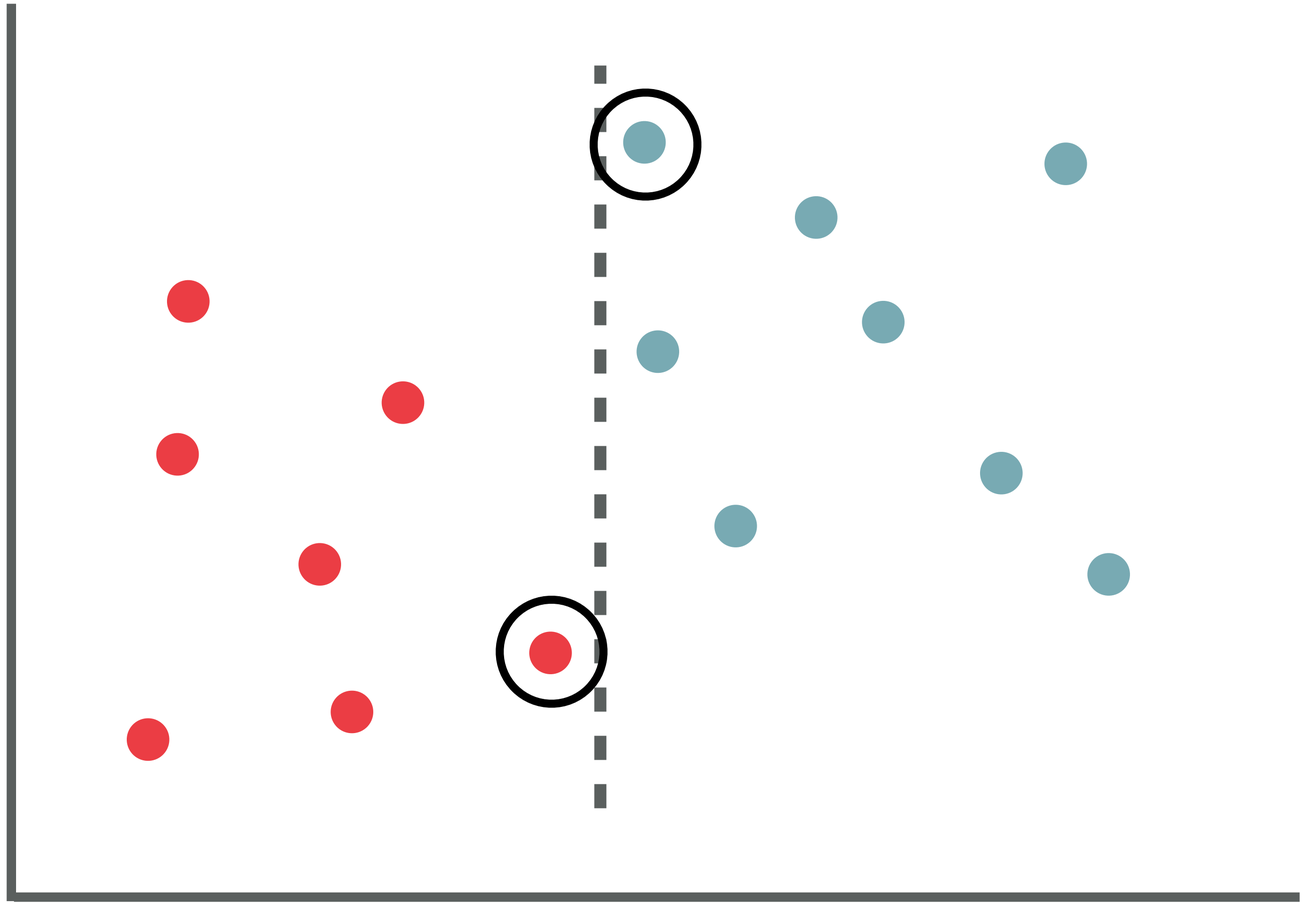
This is not the optimal separator

SVM



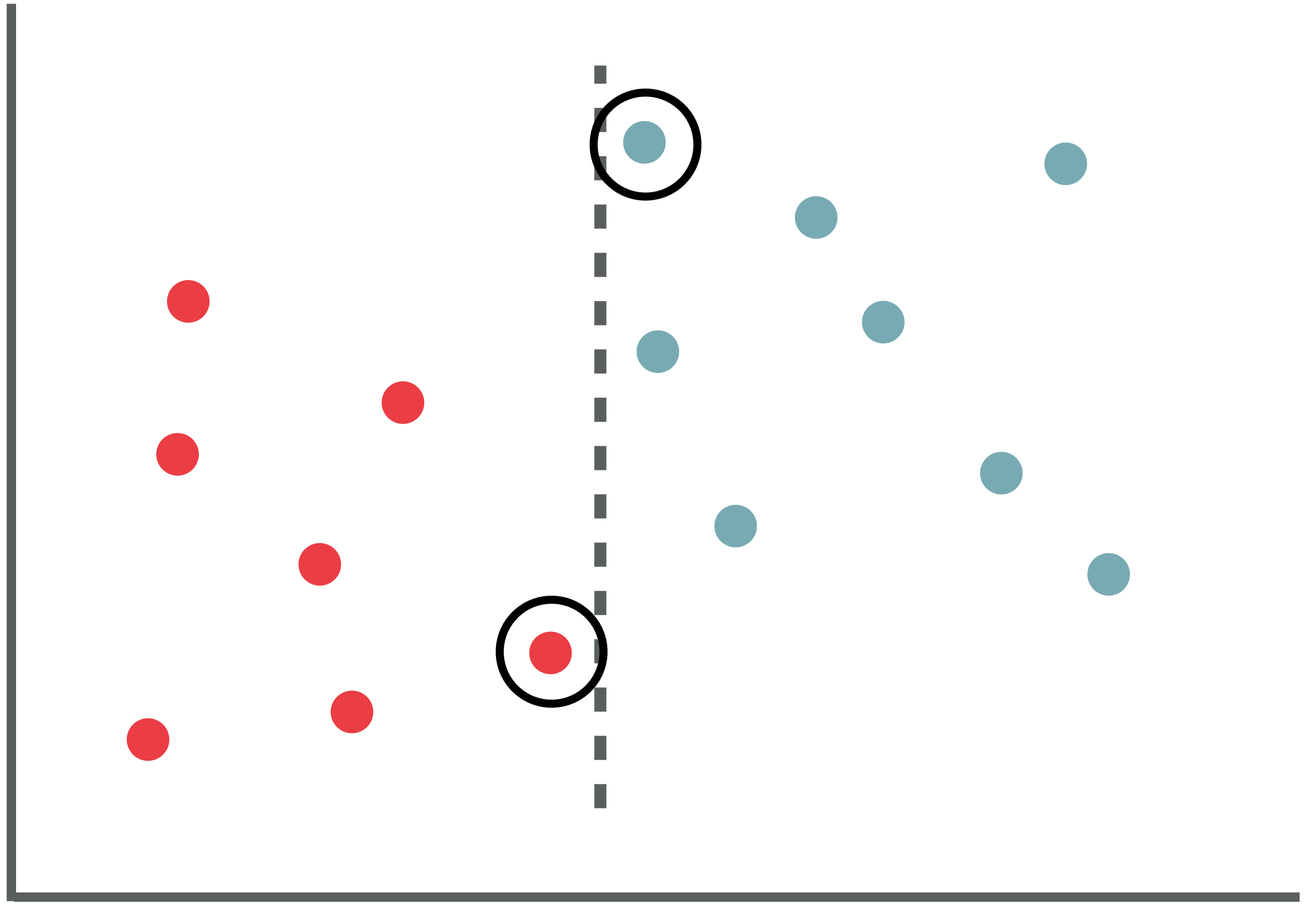
How do we find the "optimal"?

SVM



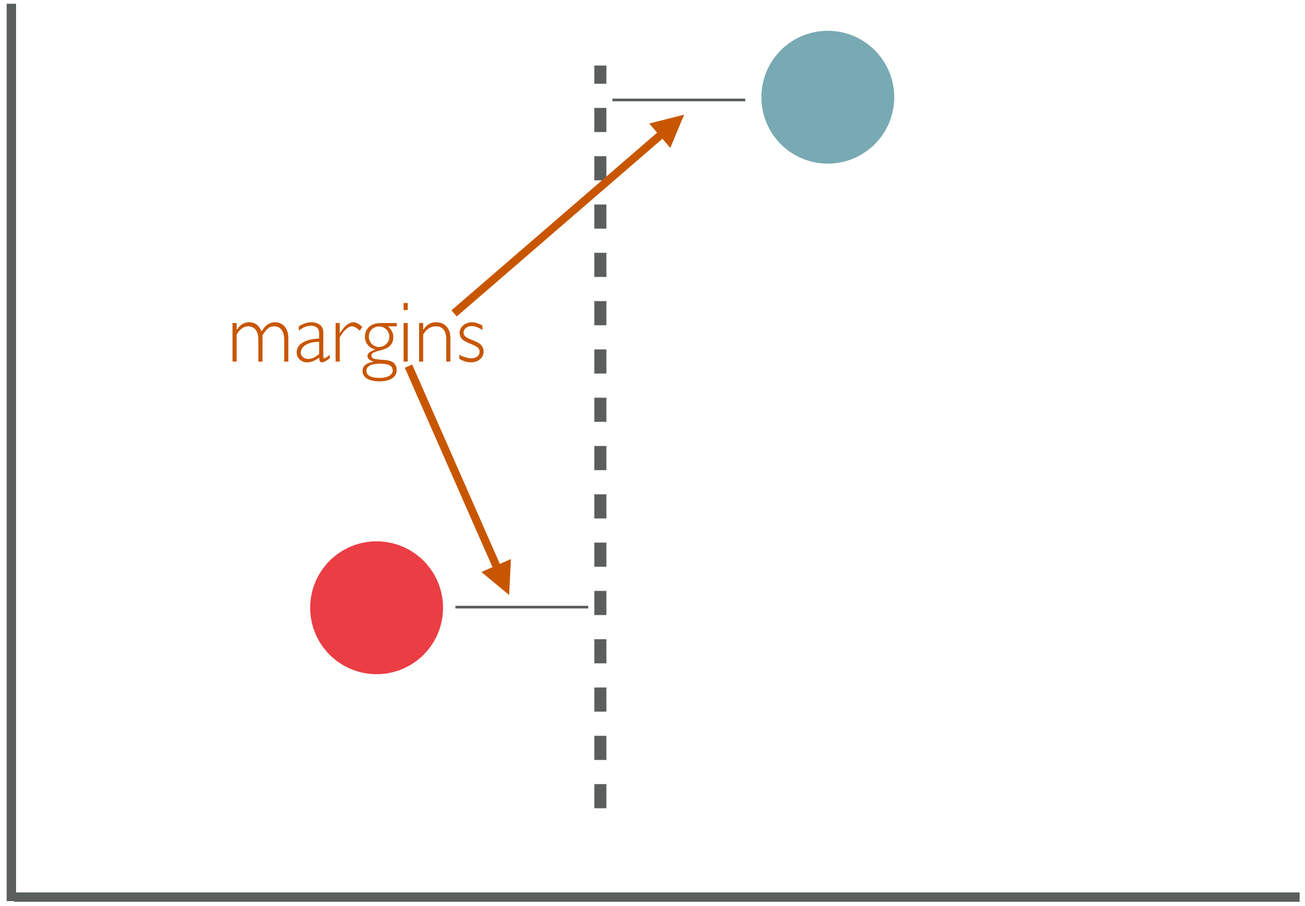
These are your
"support vectors"

SVM



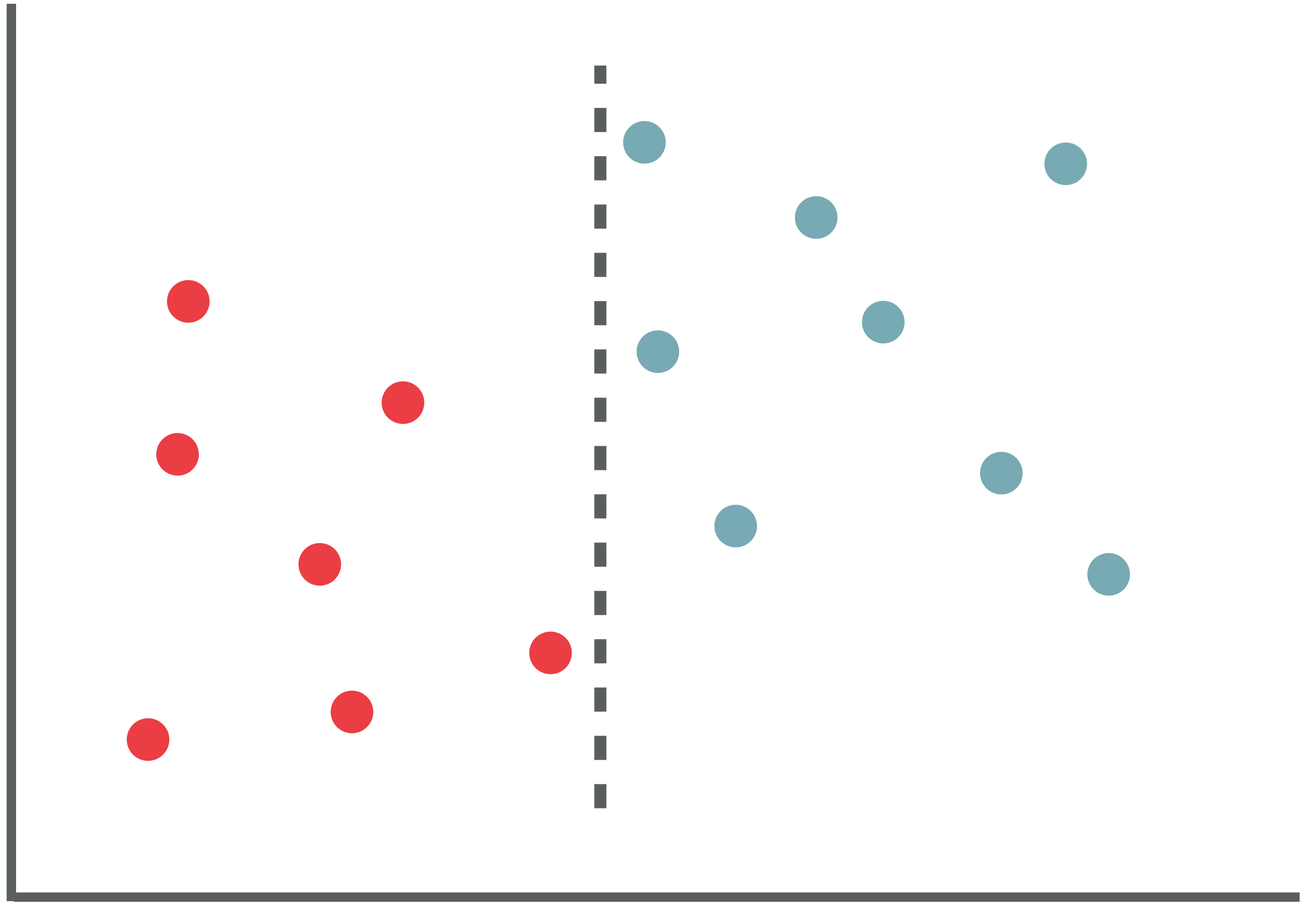
Support vectors are the points nearest to the plane

SVM



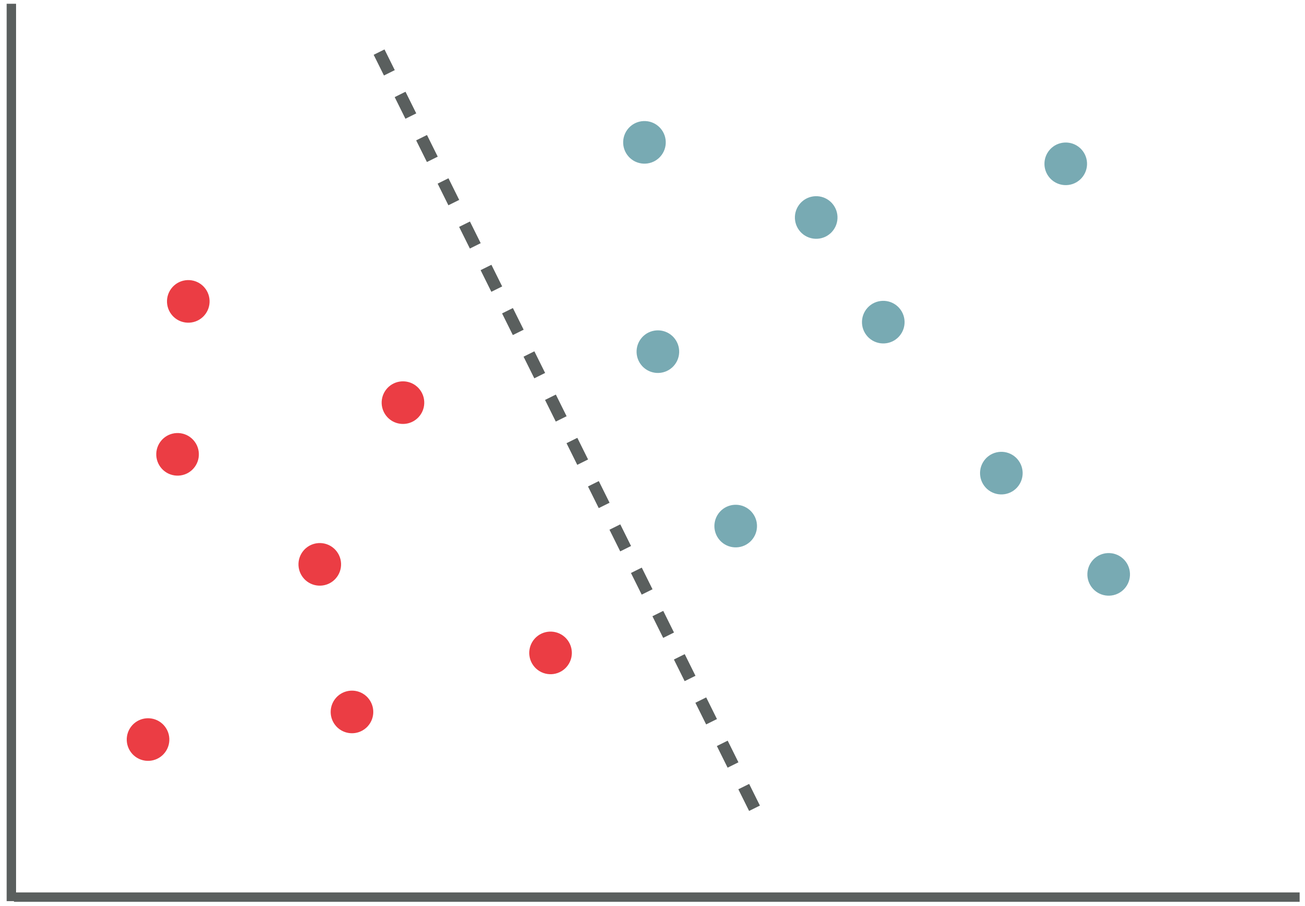
These are your margins

SVM



Goal: maximize the margins (m) such that the decision boundary is as far away from the data of both classes as possible

SVM



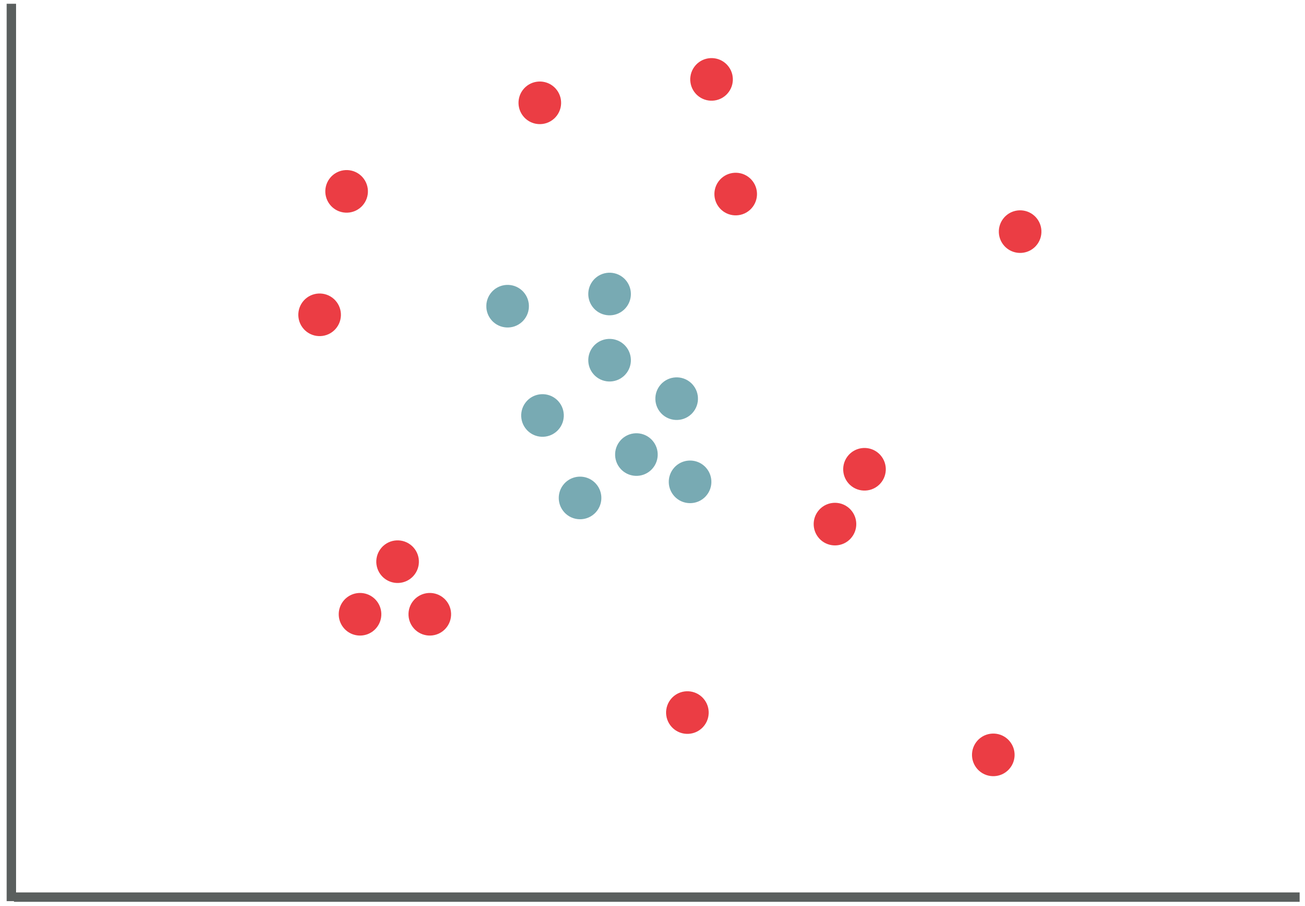
Classify!

DONE

SVM

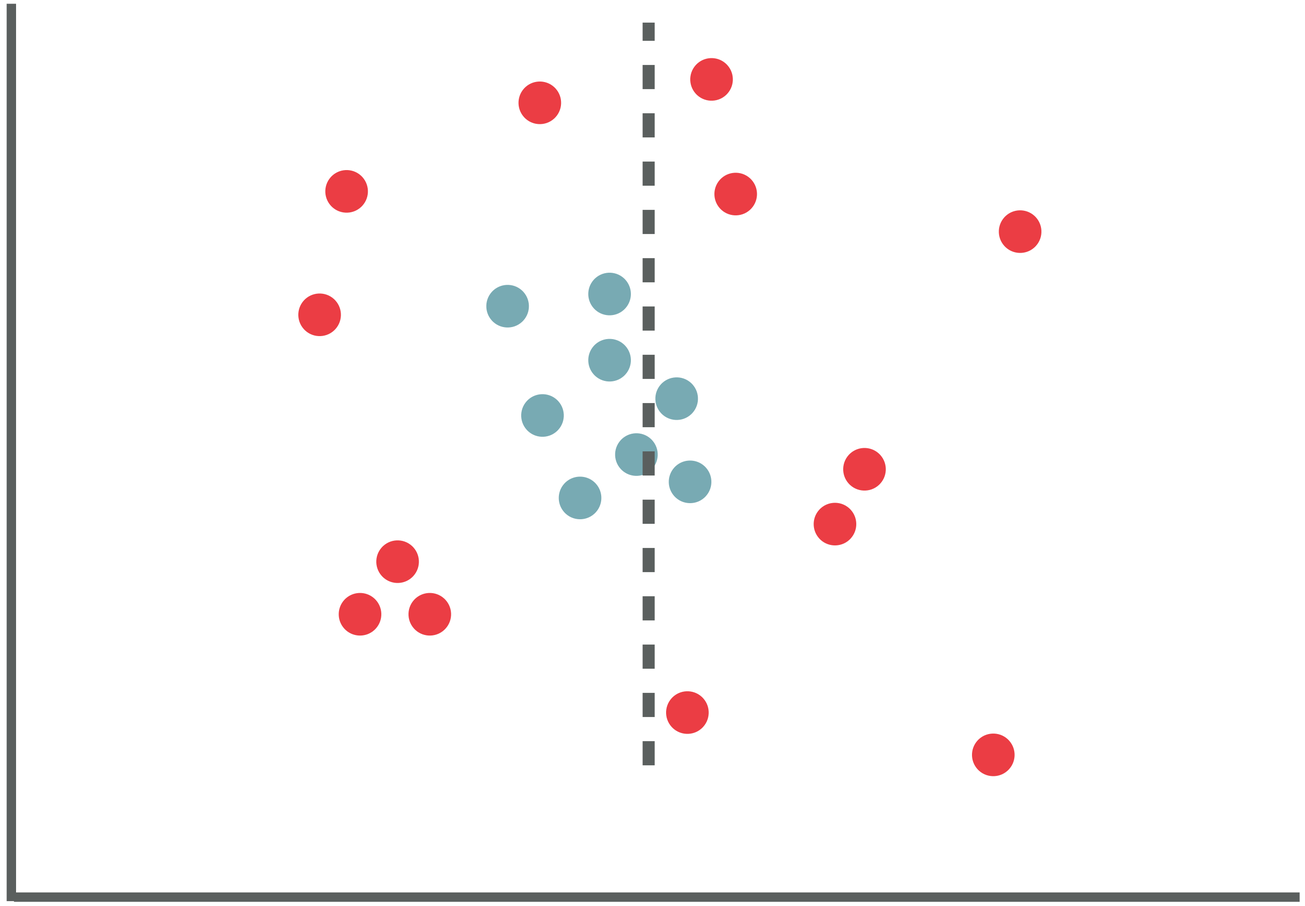
Okay… but…

SVM



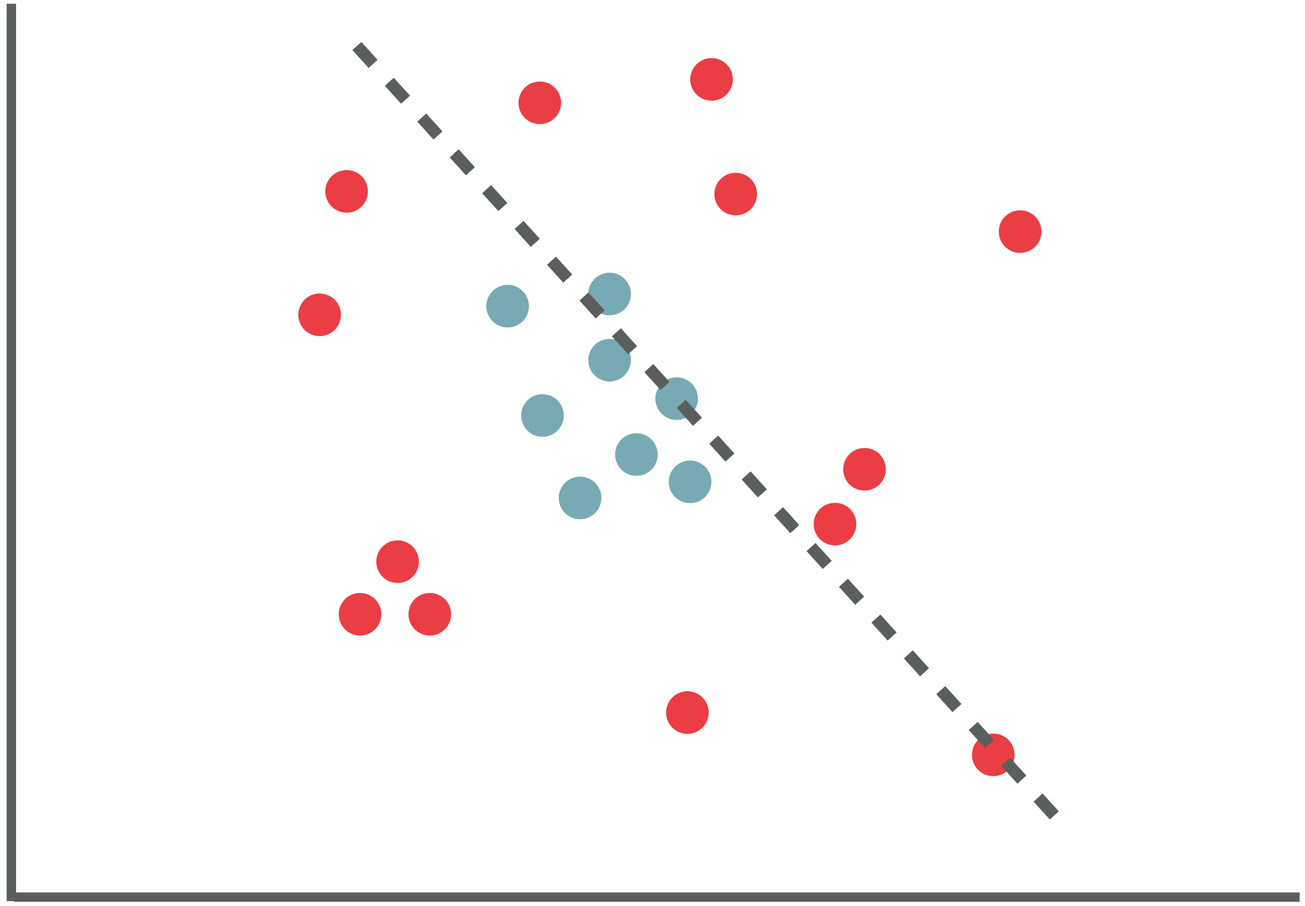
Classify!

SVM



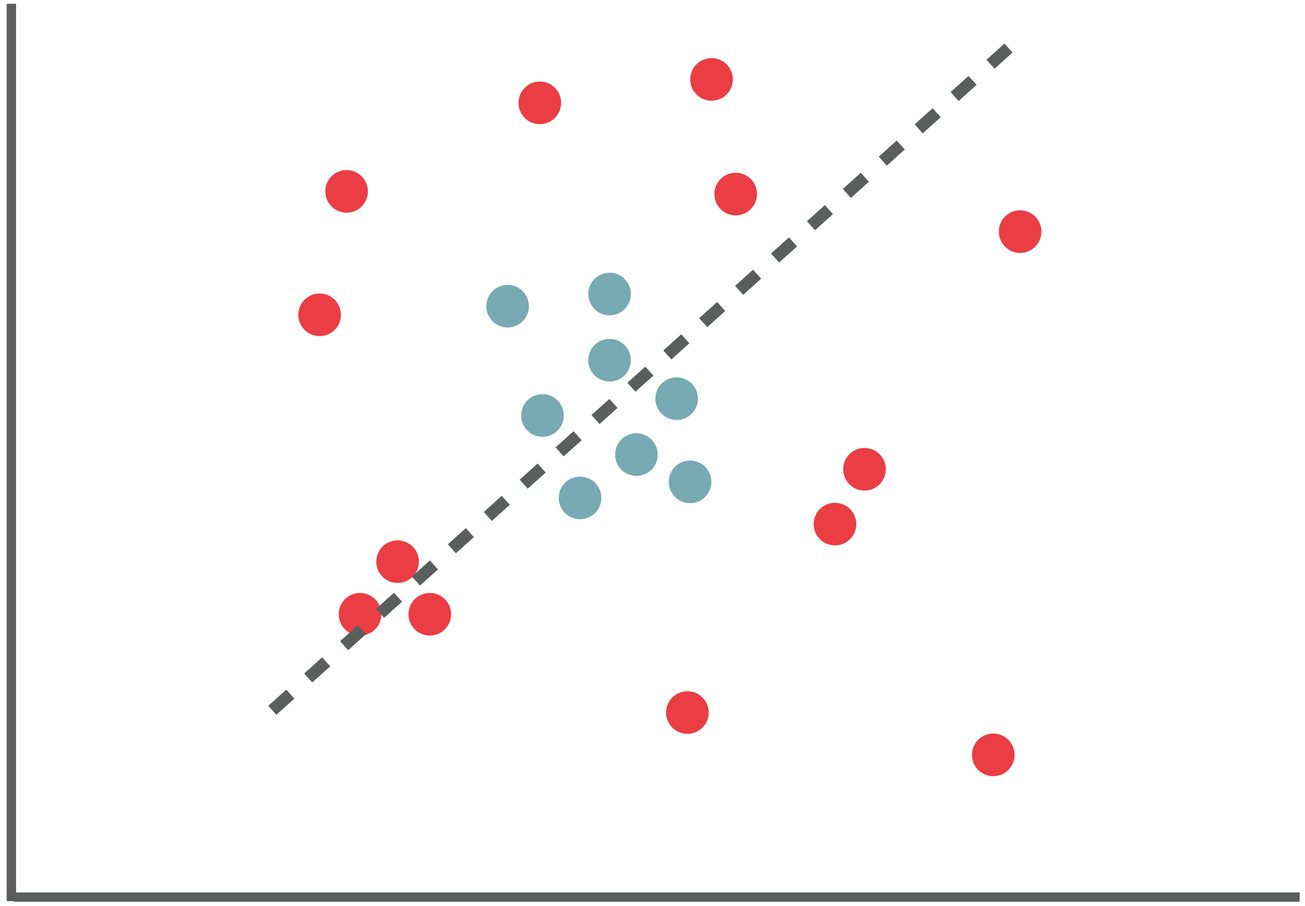
Classify!

SVM



Classify!

SVM

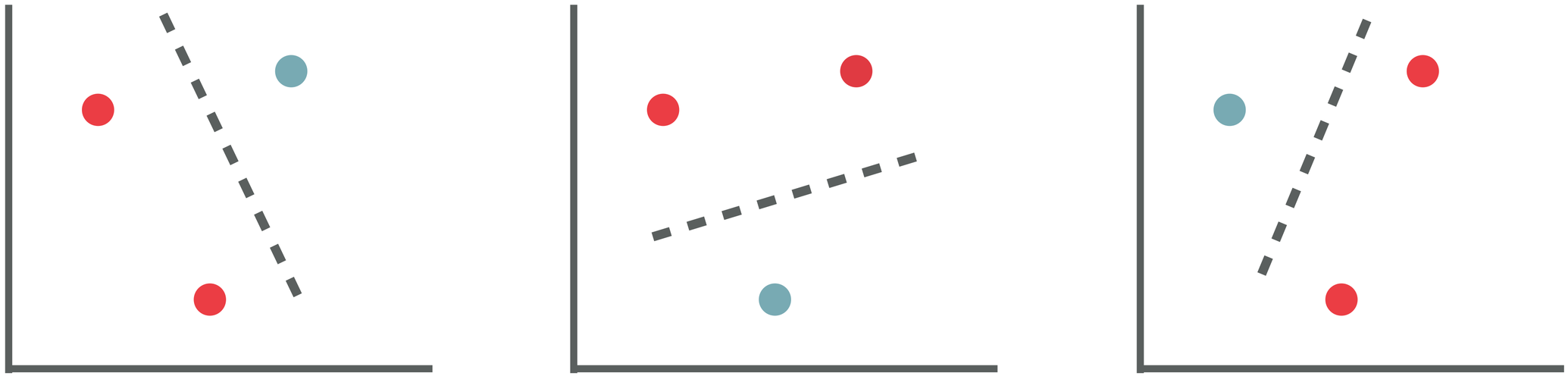


Classify!

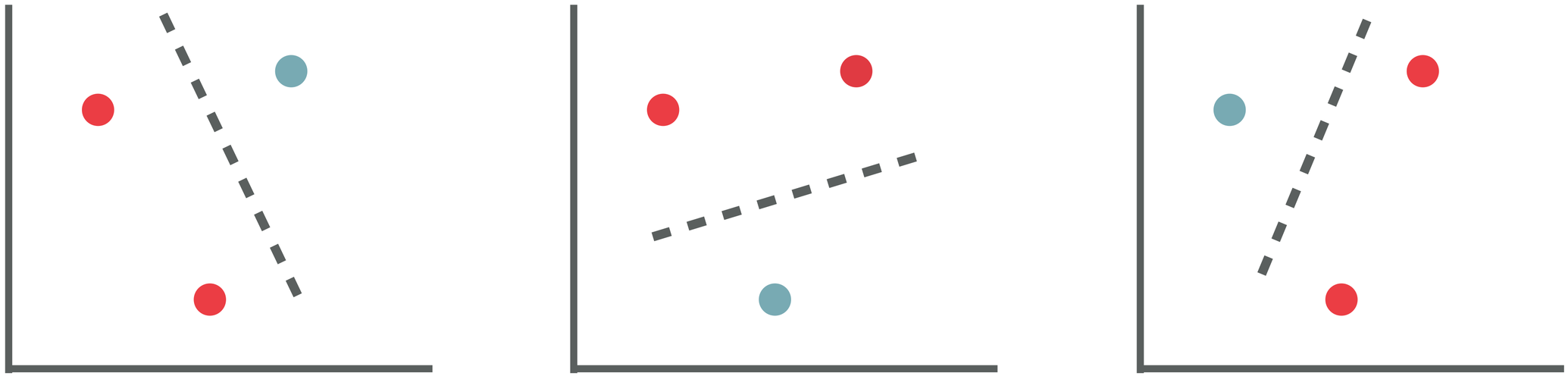
SVM

fy!

SVM



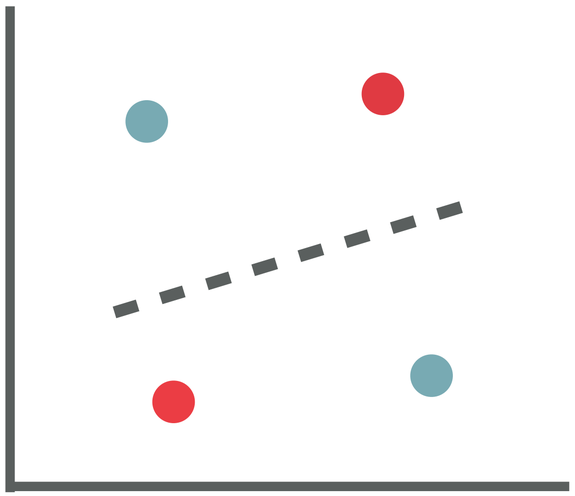
SVM



If we have 3 training points in 2D space, no matter how the points are labeled we can perfectly classify them using a linear algorithm

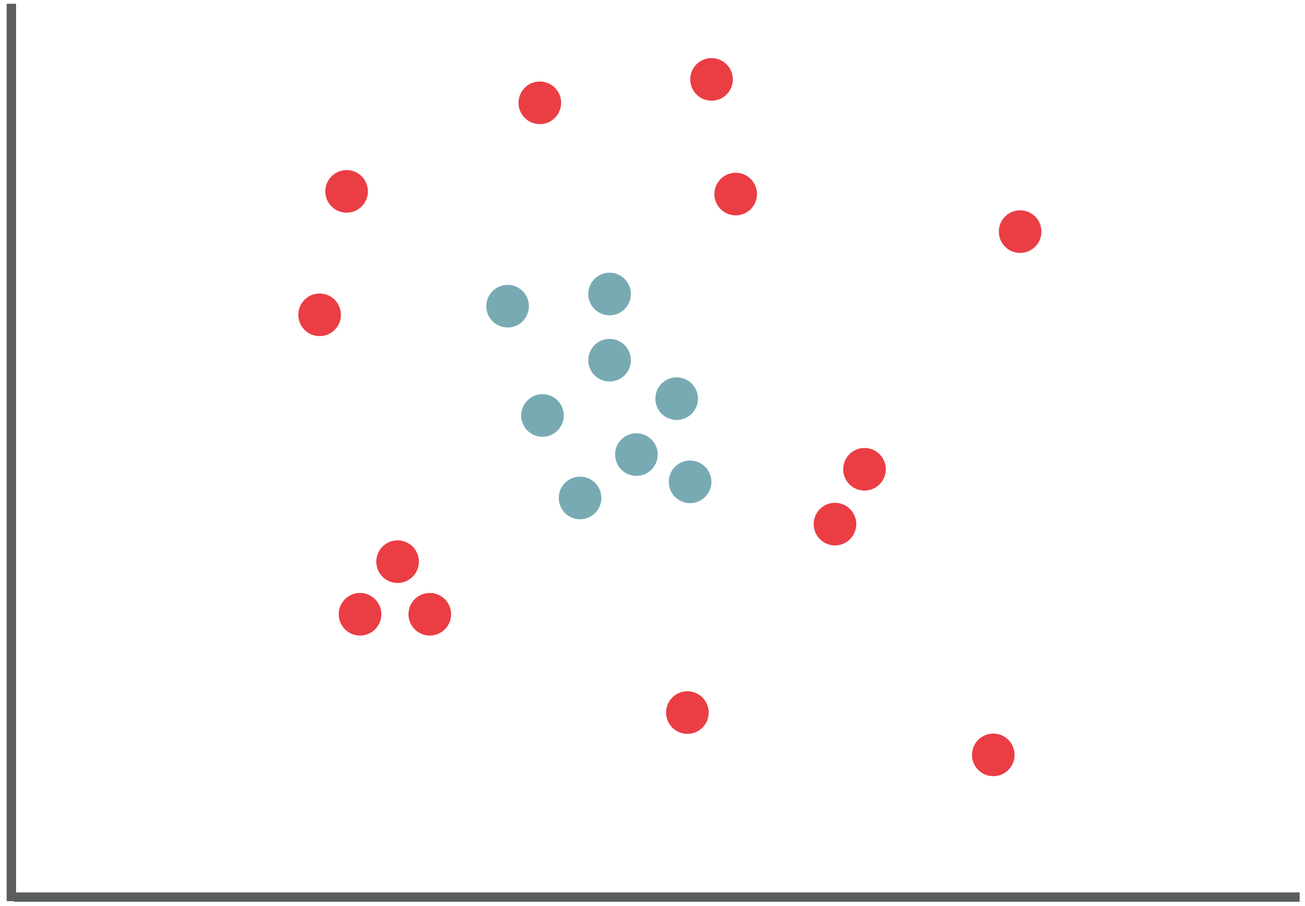
Source: Martin Law, MSU

SVM



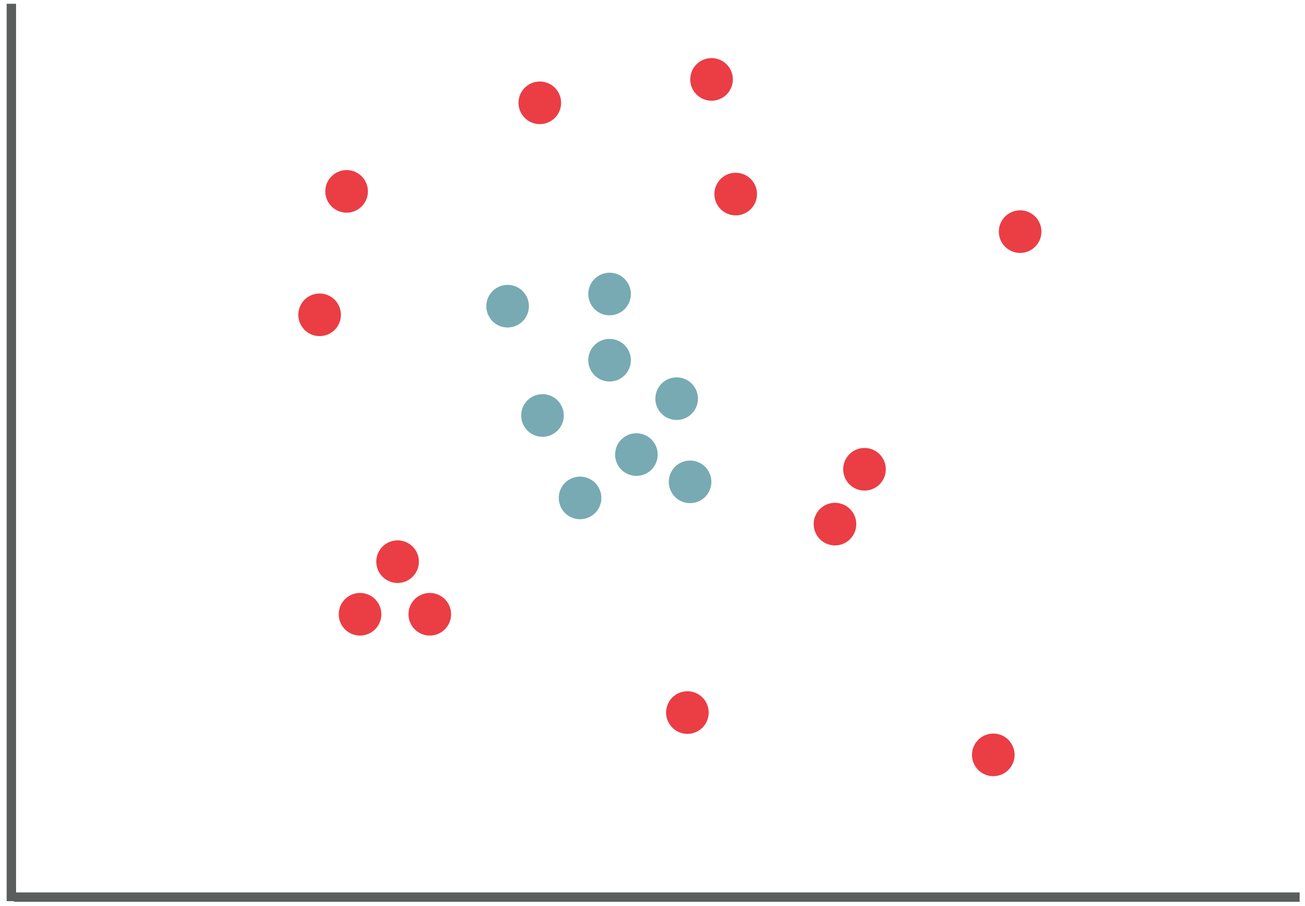
But add a fourth point, and this can fail

SVM



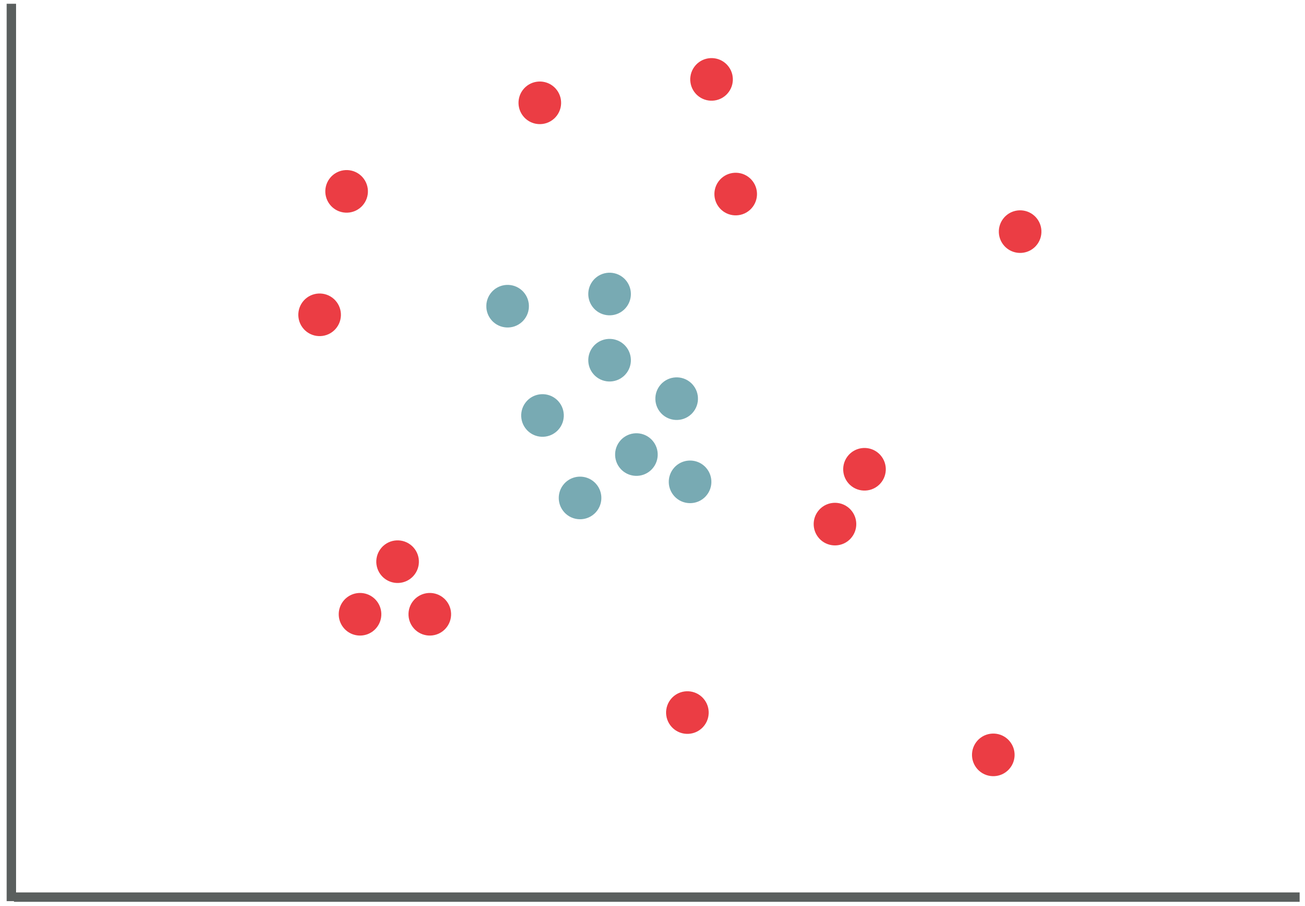
SVM: Total hack to turn data that are not optimally suited for linear classifi cation into a linearly classifi able problem by warping them into higher dimensions

SVM



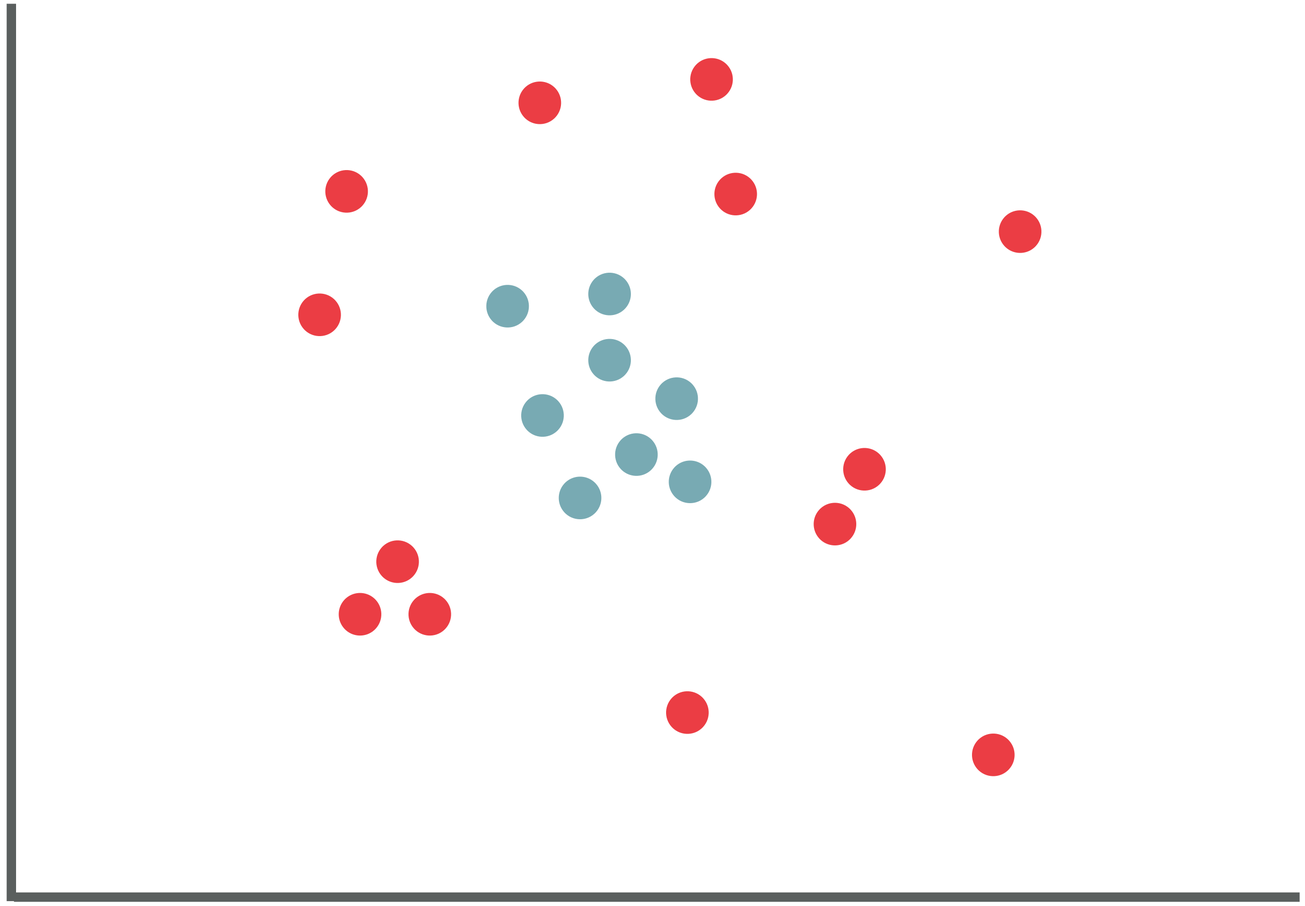
If we have *N* training points in (*N*-1)D space, no matter how the points are labeled we can perfectly classify them using a linear algorithm

SVM



But rarely do you need to go to *N*-1 space to do this

SVM



Classify!

Common SVM kernels

Examples:

$$K(\vec{x}_i, \vec{x}_j) = \vec{x}_i \cdot \vec{x}_j$$ Linear kernel

$$K(\vec{x}_i, \vec{x}_j) = \exp(-\gamma \|\vec{x}_i - \vec{x}_j\|^2)$$ Gaussian kernel

$$K(\vec{x}_i, \vec{x}_j) = \exp(-\gamma \|\vec{x}_i - \vec{x}_j\|)$$ Exponential kernel

$$K(\vec{x}_i, \vec{x}_j) = (p + \vec{x}_i \cdot \vec{x}_j)^q$$ Polynomial kernel

$$K(\vec{x}_i, \vec{x}_j) = (p + \vec{x}_i \cdot \vec{x}_j)^q \exp(-\gamma \|\vec{x}_i - \vec{x}_j\|^2)$$ Hybrid kernel

$$K(\vec{x}_i, \vec{x}_j) = \tanh(k\vec{x}_i \cdot \vec{x}_j - \delta)$$ Sigmoidal

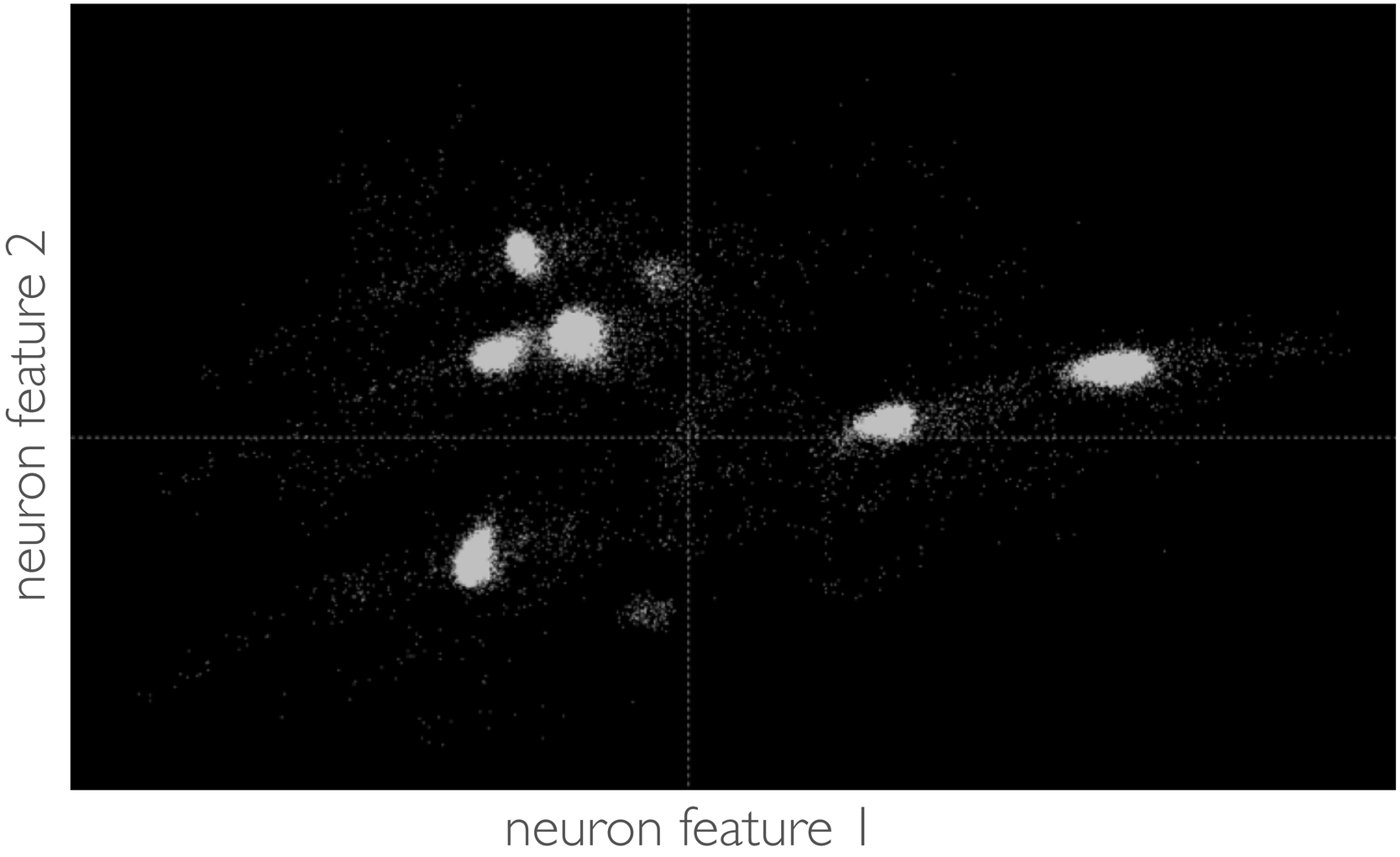
Source: Martin Law, MSU

COGS 108
Data Science in Practice

Dimensionality reduction and clustering

UC San Diego

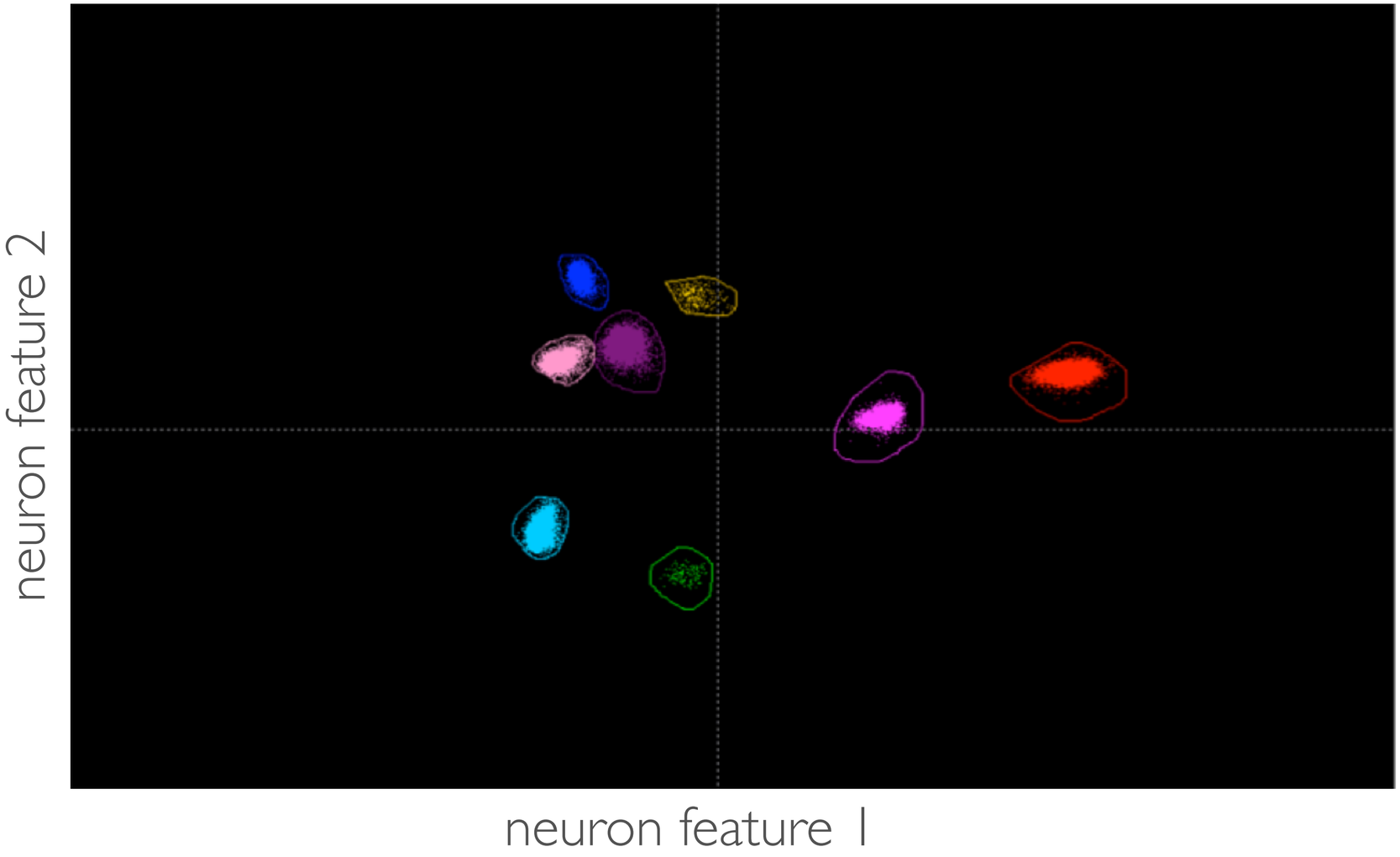
Clustering



Source: Jake Olson

UC San Diego

Clustering



Source: Jake Olson

UC San Diego

Why clustering?

- Unlabeled data – unsupervised learning
  - If it is labeled, we'll use classification!
- Want to find groups!

Source: Jake Olson

Why clustering?

- Data Compression – We could want to reduce the dimensionality in the data.
    - Dimensionality – An aspect or a feature of something is a dimension.

Source: Jake Olson

How do we cluster?

- A lot of different ways we could, but we will start with K-Means.

Source: Jake Olson

k-means

- A simple but effective clustering algorithm
- Partitions the data in to K disjoint sets (clusters)
  - disjoint - no overlap
- We then run an iterative batch algorithm
  - batch – to do to all of the items (batch of cookies)

Source: Jake Olson

k-means

- Start – Initialize a guess of k centers/"means"
  - notation – *j* simply refers to which of the *k* centers.

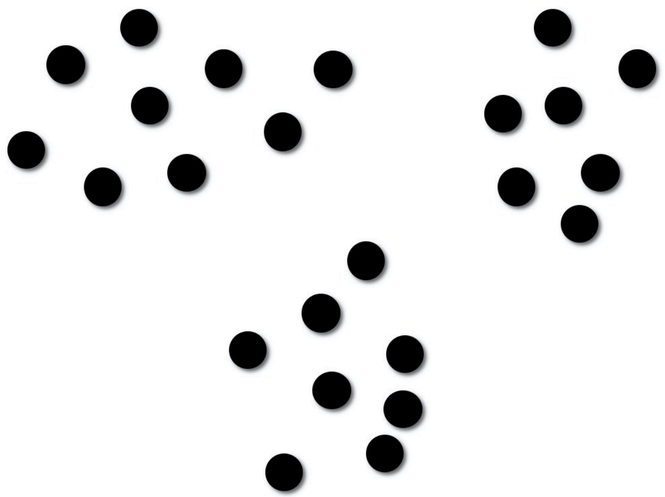
- $S^{(j)}$ is all points closest to $\boldsymbol{\mu}(j)$
  - So each "mean" gets all of the points close to it in its set.

- Update the means: $$\mu^{(j)} = \frac{1}{N_j} \sum_{n \in S^{(j)}} \mathbf{x}^{(n)}$$

- Continue until there is no change in the means or you decide to stop.

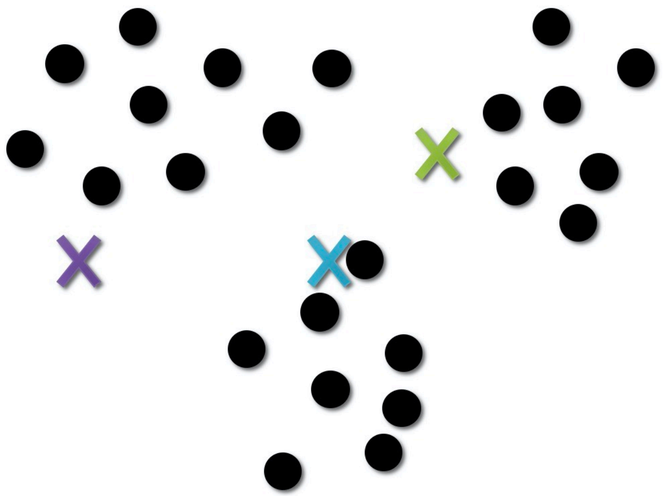
Source: Jake Olson

Initialize means



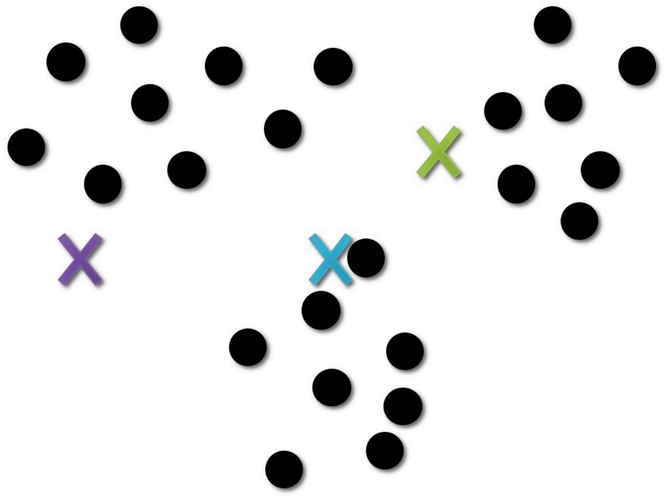
Source: Jake Olson

Initialize means



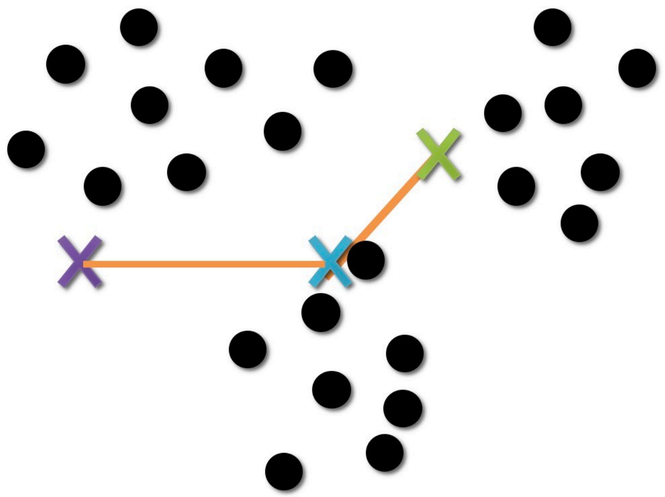
Source: Jake Olson

Label points



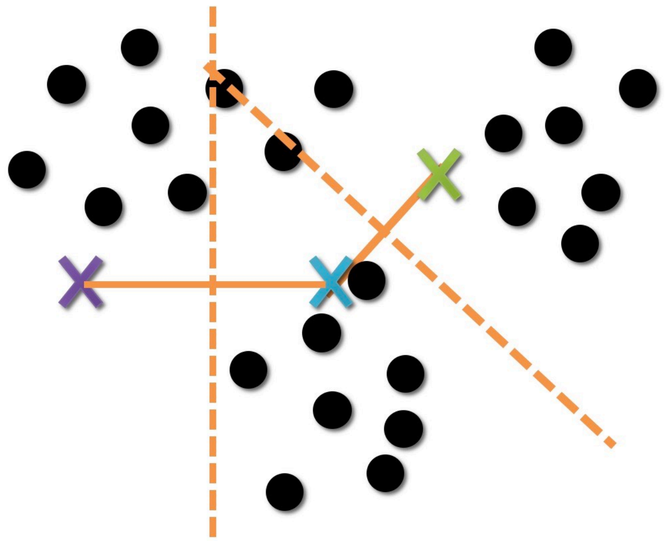
Source: Jake Olson

Connect adjacent means



Source: Jake Olson

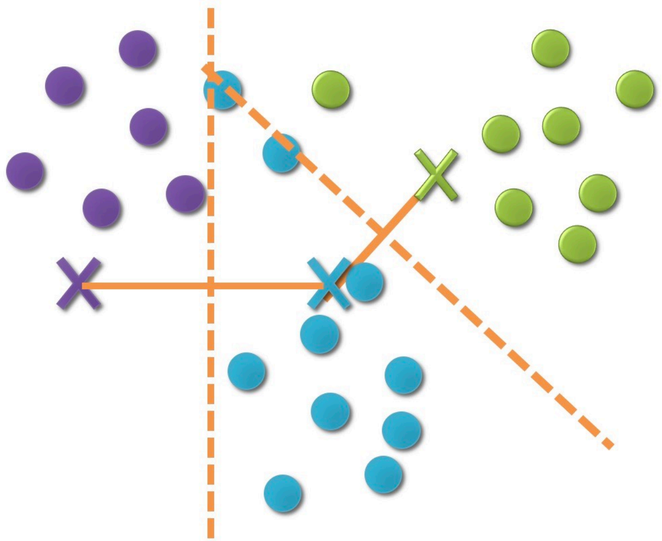
Bisect



Source: Jake Olson

UC San Diego

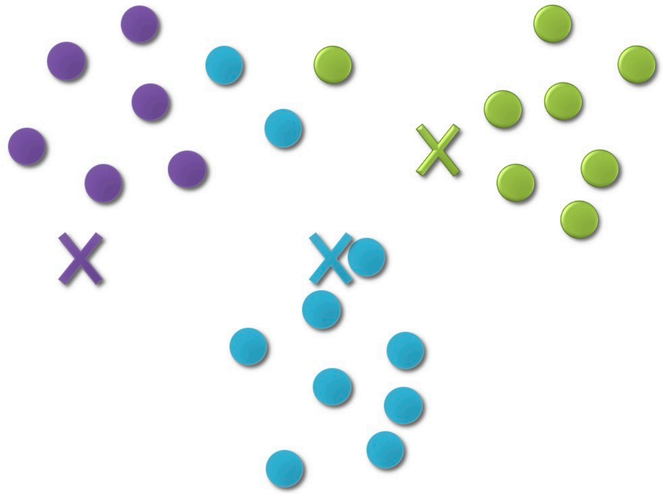
Label points



Source: Jake Olson

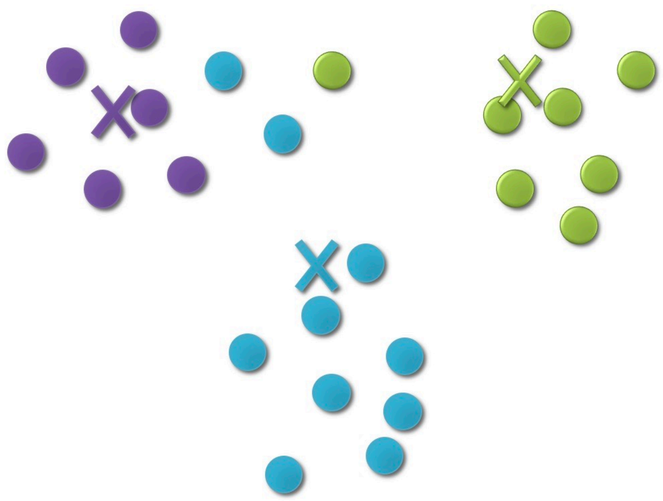
UC San Diego

Move means to centroids



Source: Jake Olson

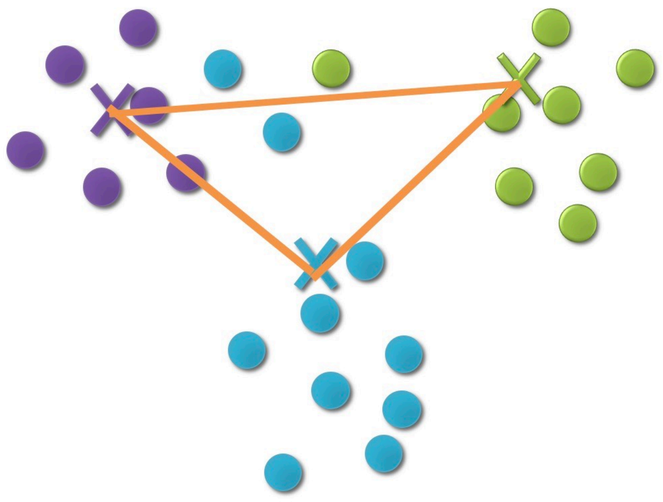
Move means to centroids



Source: Jake Olson

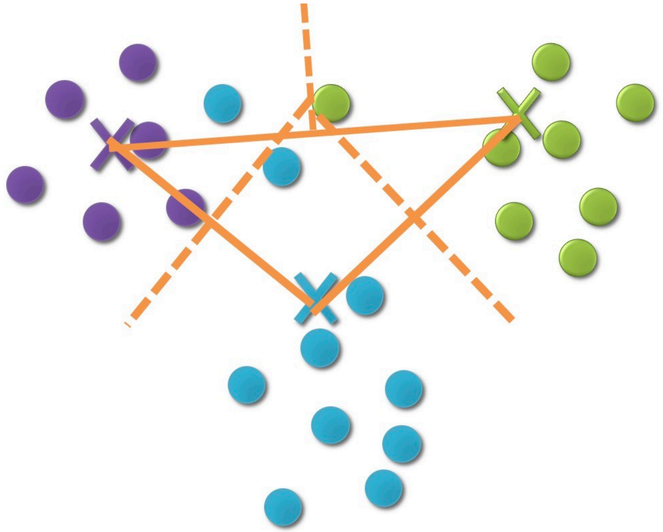
UC San Diego

If means moved, repeat!



Source: Jake Olson

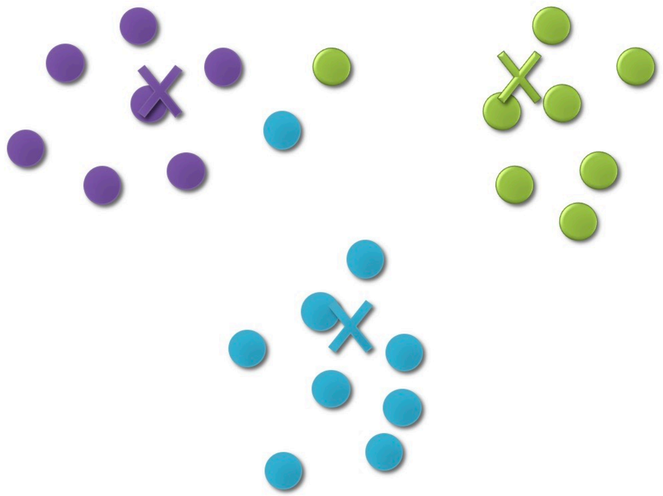
Bisect



Source: Jake Olson

UC San Diego

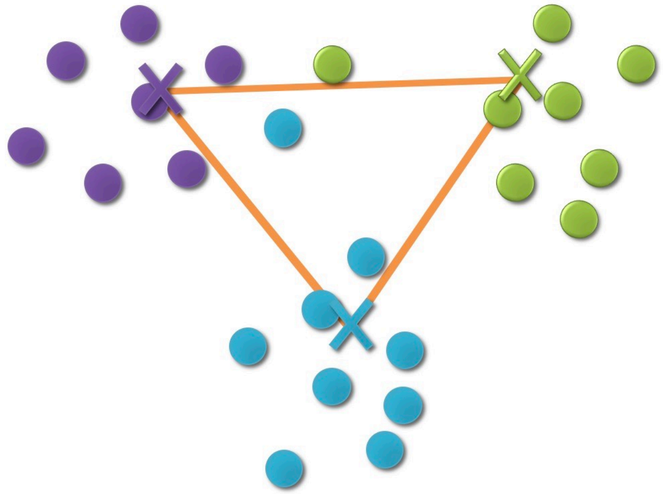
Label points



Source: Jake Olson

UC San Diego

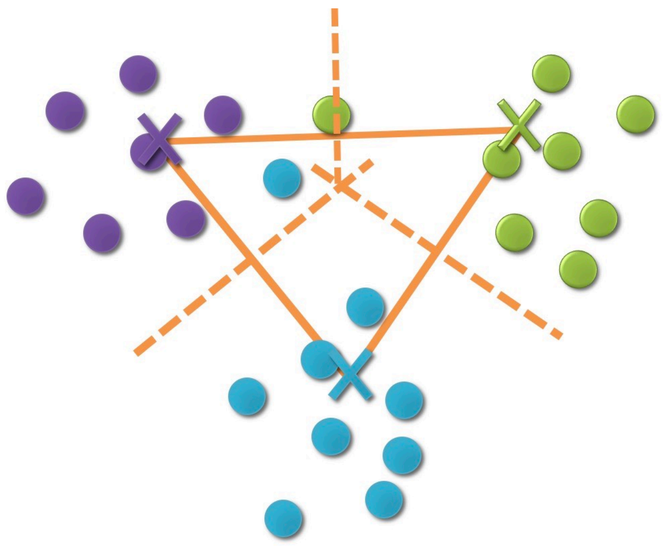
Connect adjacent



Source: Jake Olson

UC San Diego

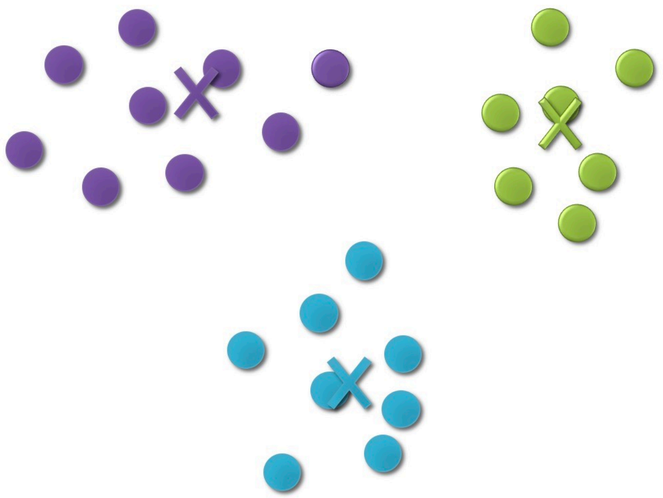
Bisect



Source: Jake Olson

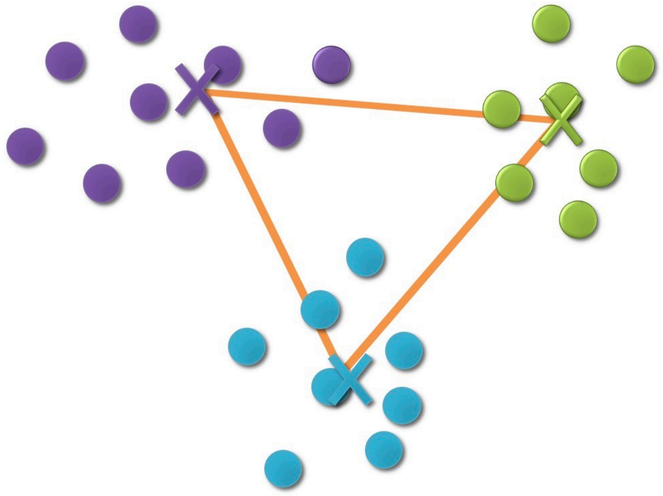
UC San Diego

Label points



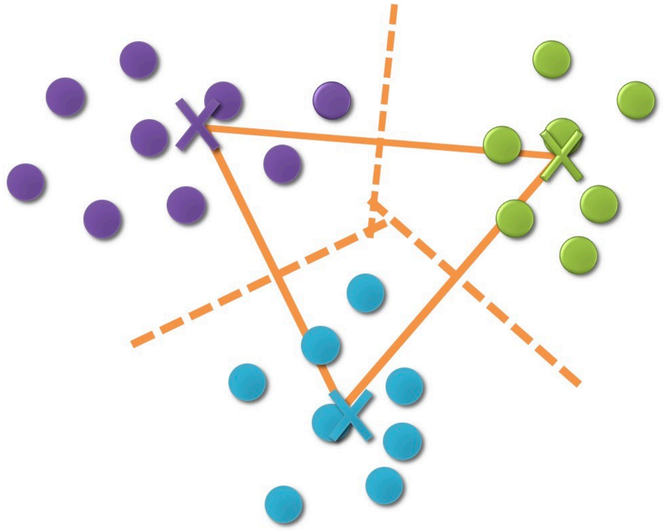
Source: Jake Olson

Connect adjacent



Source: Jake Olson

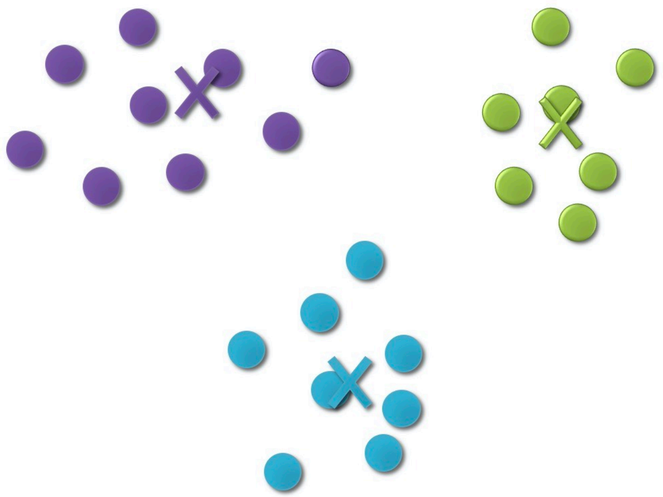
Bisect



Source: Jake Olson

UC San Diego

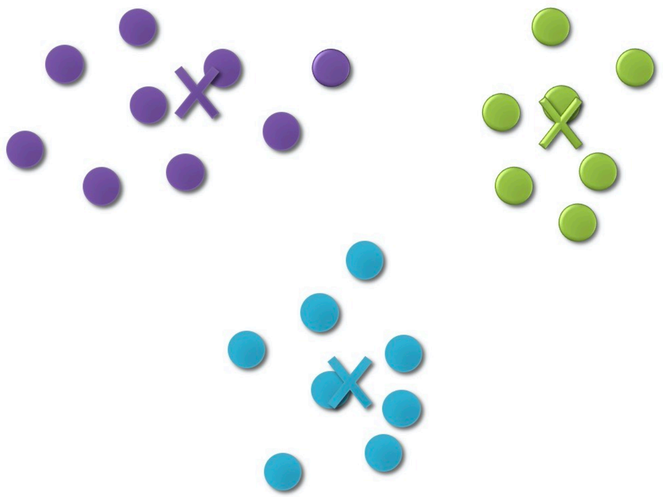
Label points



Source: Jake Olson

UC San Diego

Finished!



Source: Jake Olson

UC San Diego

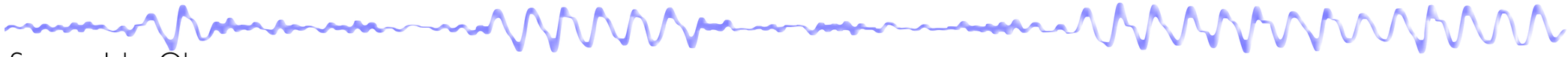
When do we reduce dimensionality?

Alicia is researching the resistance of plants in different soil to bugs. She has 10 each of 5 different types of plants. She has counted the number of bugs on each once per week for 10 weeks. Should Alicia analyze her bug count data with a clustering analysis? Why?

Source: Jake Olson

When do we reduce dimensionality?

Pat is studying cell types and morhpologies in the basal forebrain of mice. They recorded the waveform amplitude, shapes, and firing patterns of 400 neurons. Should Pat run a clustering analysis?



Source: Jake Olson

UC San Diego

When do we reduce dimensionality?

Linh is working at Apple. They have data on the accounts for all of their users, including what programs they use, how often, and for how long. They've been tasked with mining the data to find patterns that could affect marketing strategies. Should Linh run a clustering analysis?

Source: Jake Olson

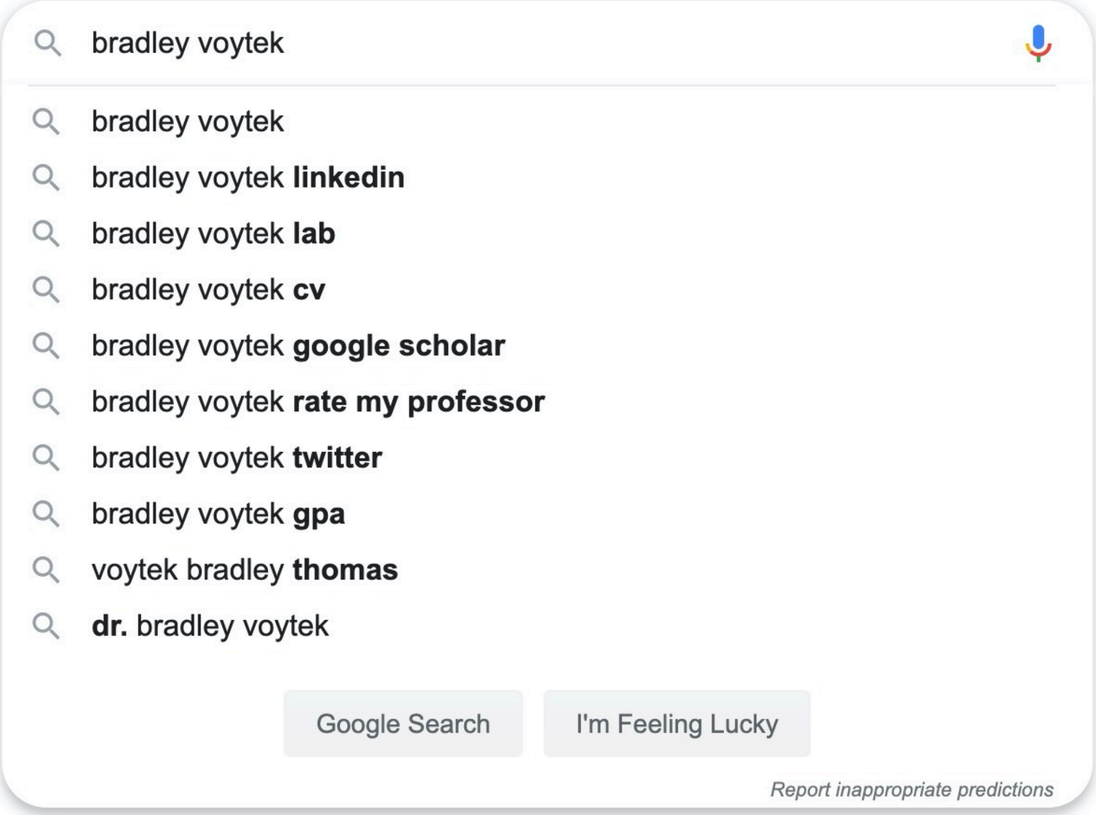
COGS 108
Data Science in Practice

Text mining and NLP

UC San Diego

What does text mining get us?

Google



UC San Diego

What does text mining get us?

Google

neuroscientists are

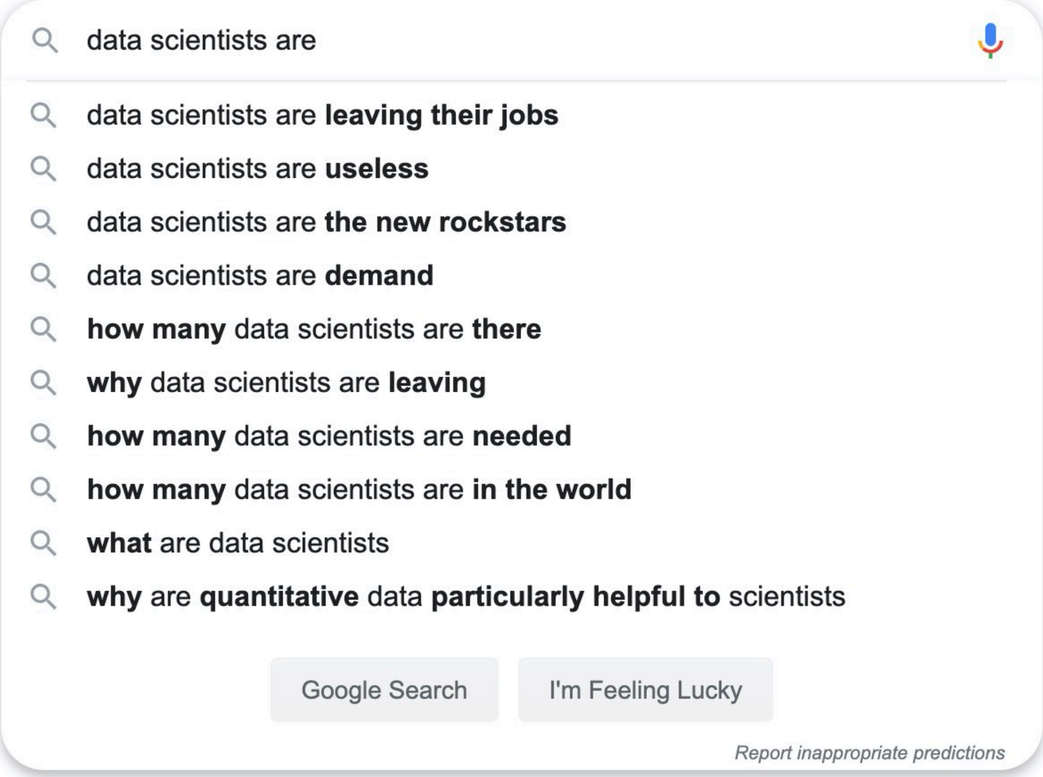
- neuroscientists are **excited by technological progress that facilitates**
- neuroscientists are **excited by technological**
- neuroscientists are **excited by technological progress**
- are neuroscientists **doctors**
- are neuroscientists **in demand**
- are neuroscientists **smart**
- are neuroscientists **medical doctors**
- are neuroscientists **well paid**
- **why** neuroscientists are **important**
- **in general** neuroscientists are **interested in the**

Google Search | I'm Feeling Lucky

Report inappropriate predictions

What does text mining get us?

Google



Simple word visualization

Simple word visualization

cognition
cognitive
language
learning
interaction
embodied
brain
human
design
attention
computer
data
memory
social
processing
mind
perception
visual
cortex
neural
event
interfaces
big
thinking
distributed
machine
spatial
development
ethnography
gesture
ways
aging
children
vision
speak
society
theory
linguistics
informatics
biological
comprehension
creative
world
fostering
face
online
control
network
eye
mental
use
basis
more
think
active
tools
abstract
imaging
recognition
abnormal
amygdala
factors
create
modeling
shape
mining
parent
premotor
imagine
normal
neuroscience
decision
multimodal
pragmatics
infant
parietal
associative
people
between
neuro
mathematical
learn
knowledge
word
literal
genetic
hippocampus
data
excellence
science
music
understanding
social
brain
tracking
episodic
shape
aging
tools
spatial
things
developing
interactions
languages
children
empowering
distributed
example
making
relationship
meaning
behavior
world
society
people
think
working
learn
face
ways
semantics
simulation
medical
experimental
automated
information
cortex
integration
cohesive
learning
monoamines
multimodal
oscillatory
active
create
genetic
normal
abstract
relationships
design
event
factors
bayesian
machine
learning
factors
online
psycholinguistics
neuro
abstraction
mining
medical
creative
neuromodulation
cognitive
neuroimaging
inference
computer
embodied
abnormal
computing
developing
interaction
integration
modeling
music
comprehension
world
creative
neural
motion
gesture
memory
shape
episodic
aging
tracking
parietal
biological
individual
robotics
visual
genetic
network
infant
literal
language
understanding
differences
phenotypes
imaging
spatial
things
social
developing
interactions
children
drivers
cognition
behavior
meaning
thinking
working
learn
people
think
face
ways
semantics
simulation
medical
experimental
automated
information
integration
cortex
associative
pragmatics
abstract

Text basics

- cognitive science is the study of cognition
- data science is the use of data to uncover knowledge, including the cognitive

Text basics

- cognitive science is the study of cognition
- data science is the use of data to uncover knowledge, including the cognitive

[“cognitive”, “science”, ”is”, “the”, “study”, “of”, “cognition”, “data”, “use”, “to”, “uncover”, “knowledge”, “including”]

Vector space model

- cognitive science is the study of cognition
- data science is the use of data uncover knowledge, including the cognitive

[“cognitive”, “science”, ”is”, “the”, “study”, “of”, “cognition”, “data”, “use”, “uncover”, “knowledge”, “including”]

[1 1 1 1 1 1 1 0 0 0 0 0]

[1 1 1 2 0 1 0 2 1 1 1 1]

TF-IDF

$$w_{x,y} = tf_{x,y} \times \log\left(\frac{N}{df_x}\right)$$

TF-IDF

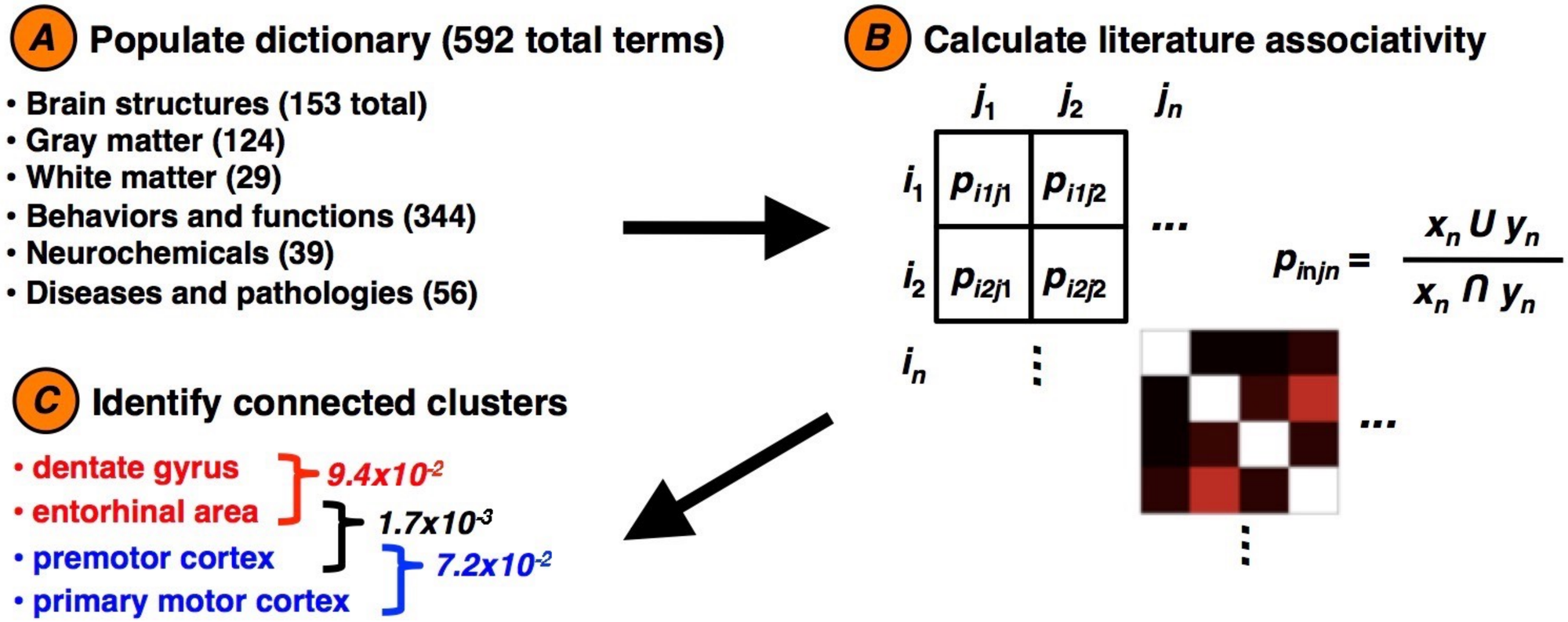
Term x within document y

$tf_{x,y}$ = frequency of x in y

df_x = number of documents containing x

N = total number of documents

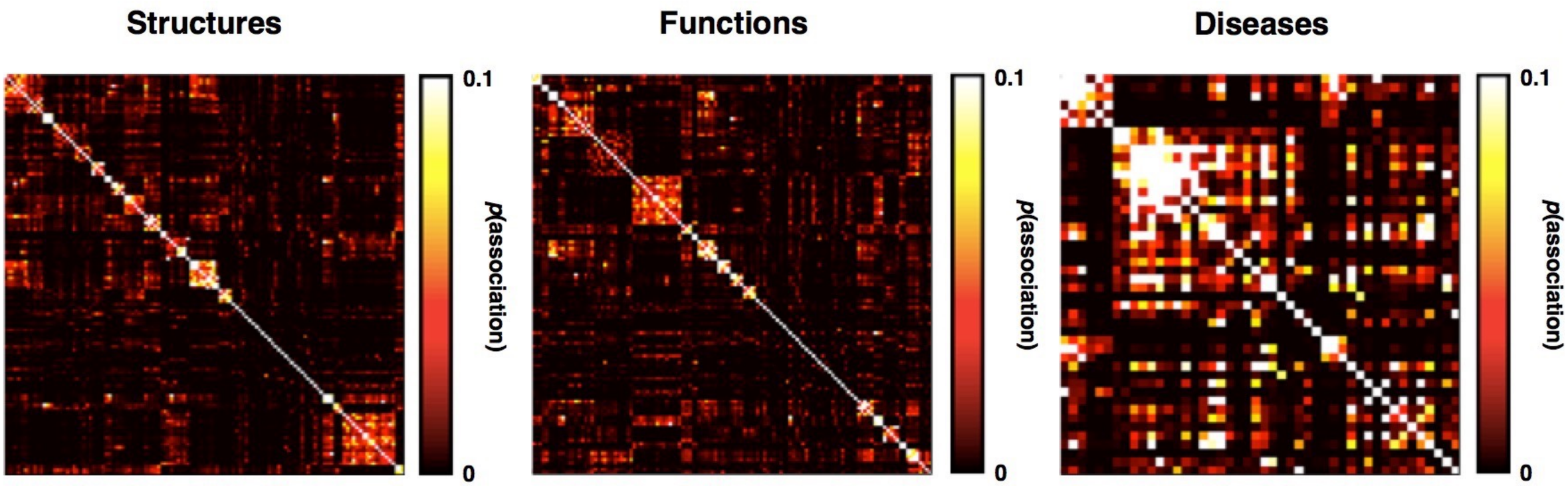
brainSCANr



Source: Voytek & Voytek, *J Neurosci Methods* 2012

UC San Diego

brainSCANr



Source: Voytek & Voytek, *J Neurosci Methods* 2012

UC San Diego

brainSCANr

Language comprehension
Language processing
Language production
Lexical processing
Lexical retrieval
Phonological encoding
Picture naming
Semantic processing
Sentence comprehension
Sentence production
Syntactic processing
Word comprehension
Word production

Anxiety
Bipolar disorder
Depression
Obsessive compulsive disorder
Panic disorder
Schizophrenia
Social phobia

Language comprehension
Sentence comprehension
Syntactic processing
Broca's aphasia
Wernicke's aphasia
Broca's area
Wernicke's area

Parkinson's
Parkinson's disease
Caudate nucleus
Globus pallidus
Putamen
Substantia nigra

Source: Voytek & Voytek, *J Neurosci Methods* 2012

UC San Diego