Lecture 1.1: Introduction
CSC 84020 - Machine Learning

Andrew Rosenberg

January 29, 2010

Today

- Introductions and Class Mechanics.

Background about me

Me:

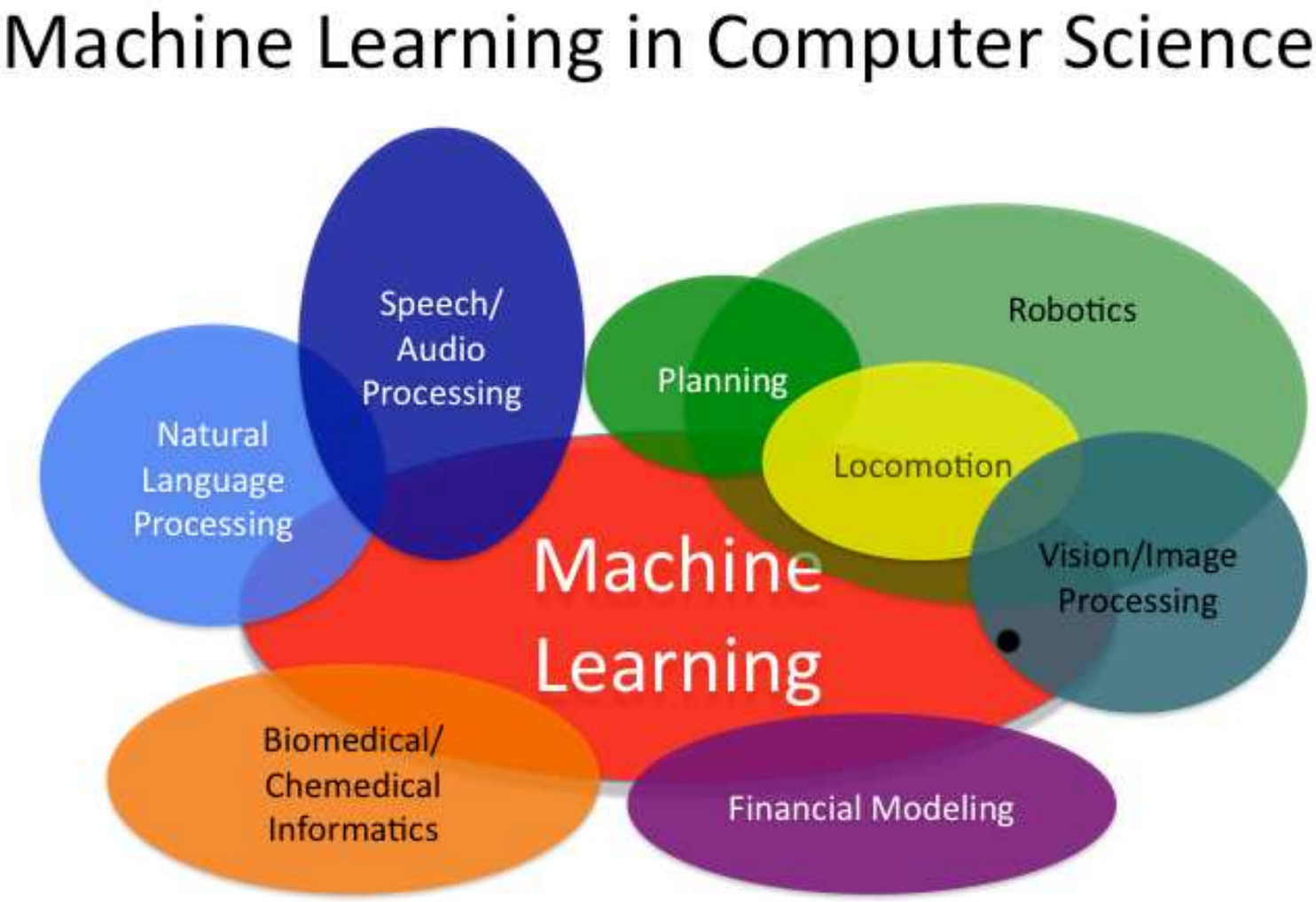
- Graduated from Columbia in 2009
- Research Speech and Natural Language Processing (Computational Linguistics)
- Specifically analyzing the intonation of speech.
- Written papers on Evaluation Measures
- All of my research has relied heavily on **Machine Learning**

Background about you

You:

- Why are you taking this class?
- What is your background in and comfort with:
  - Calculus
  - Linear Algebra
  - Probability and Statistics
- What do you hope to get from this class?

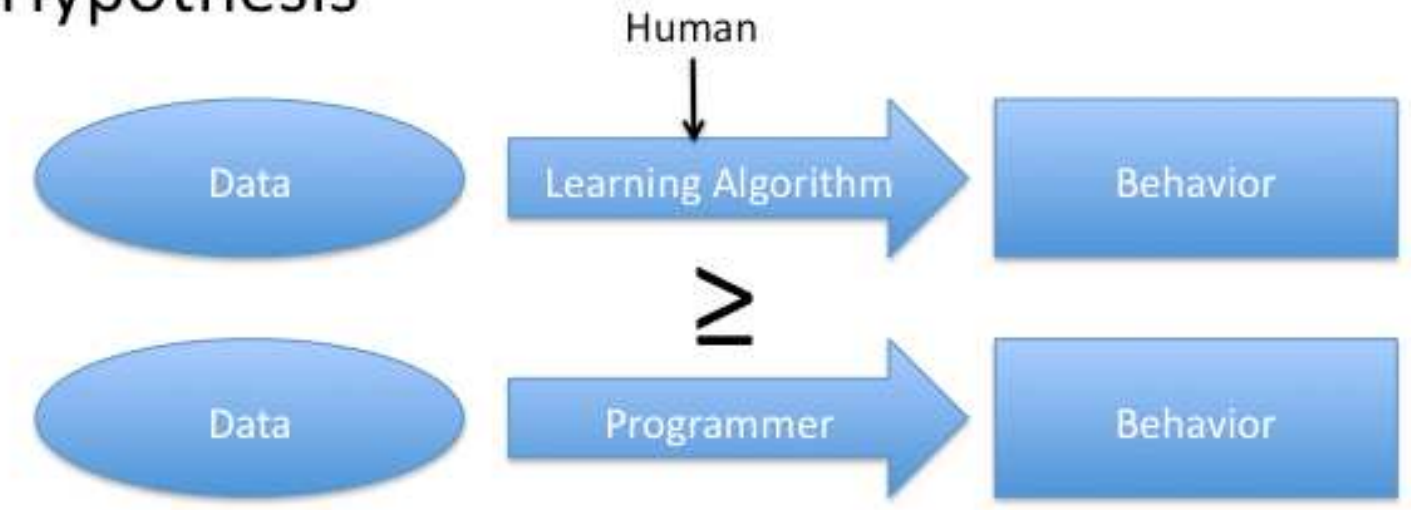
Why does anyone care about Machine Learning?



What IS Machine Learning

- Automatically identifying patterns in data
- Automatically making decisions based on data.

- Hypothesis



Major Tasks of Machine Learning

Major Tasks

- Classification
- Regression
- Clustering

Classification

- Identify which of N classes a data point belongs to.

$\mathbf{x}$ is a feature vector based on some entity x.

$$\mathbf{x} = \begin{pmatrix} f_0(x) \\ f_1(x) \\ \dots \\ f_{n-1}(x) \end{pmatrix}$$

Also, sometimes,

$$\mathbf{x} = \begin{pmatrix} x_0 \\ x_1 \\ \dots \\ x_{n-1} \end{pmatrix}$$

Target Values

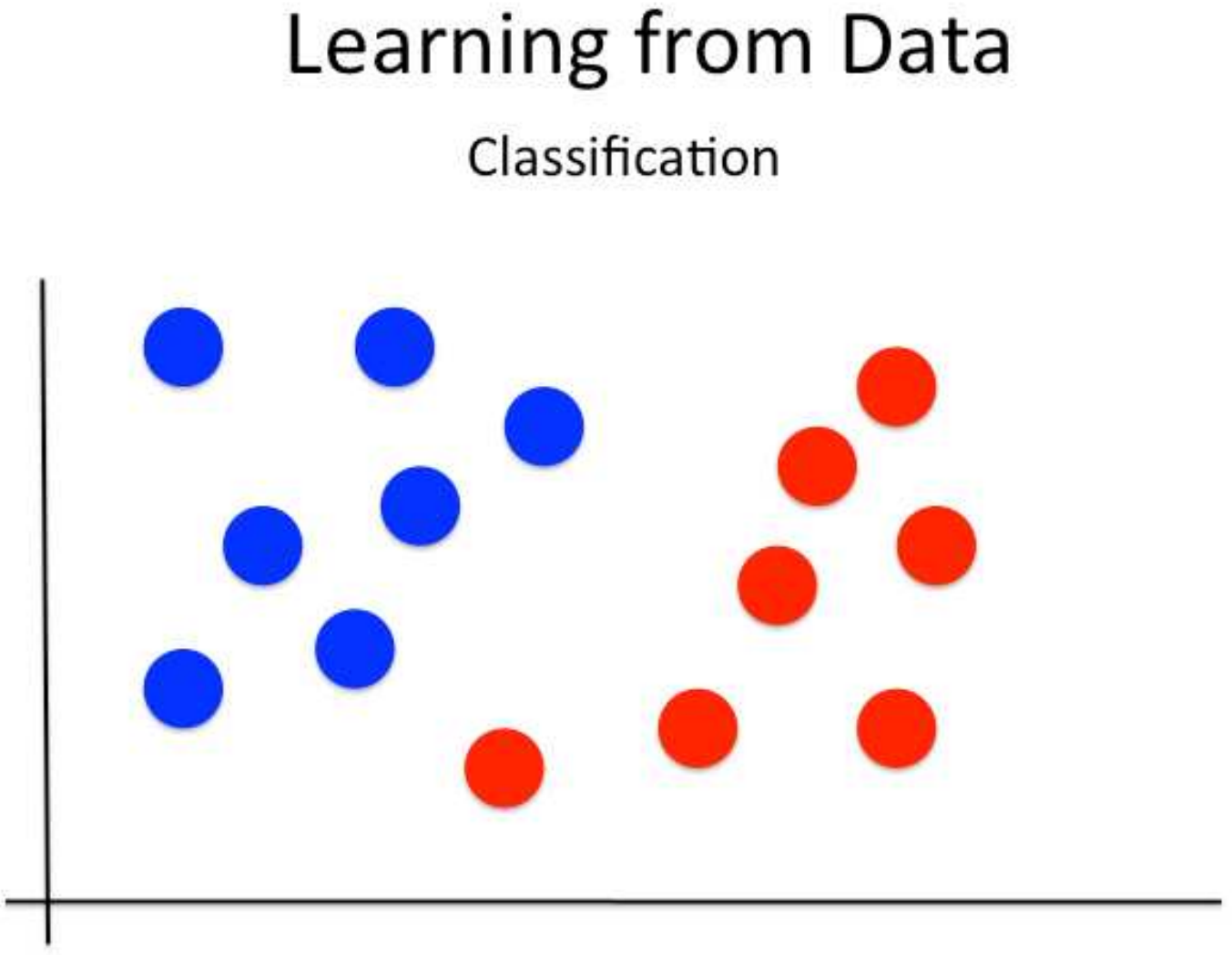
In **supervised** approaches, in addition to the data point x, we will also have some target value t.

In classification, t represents the **class** of the data point.

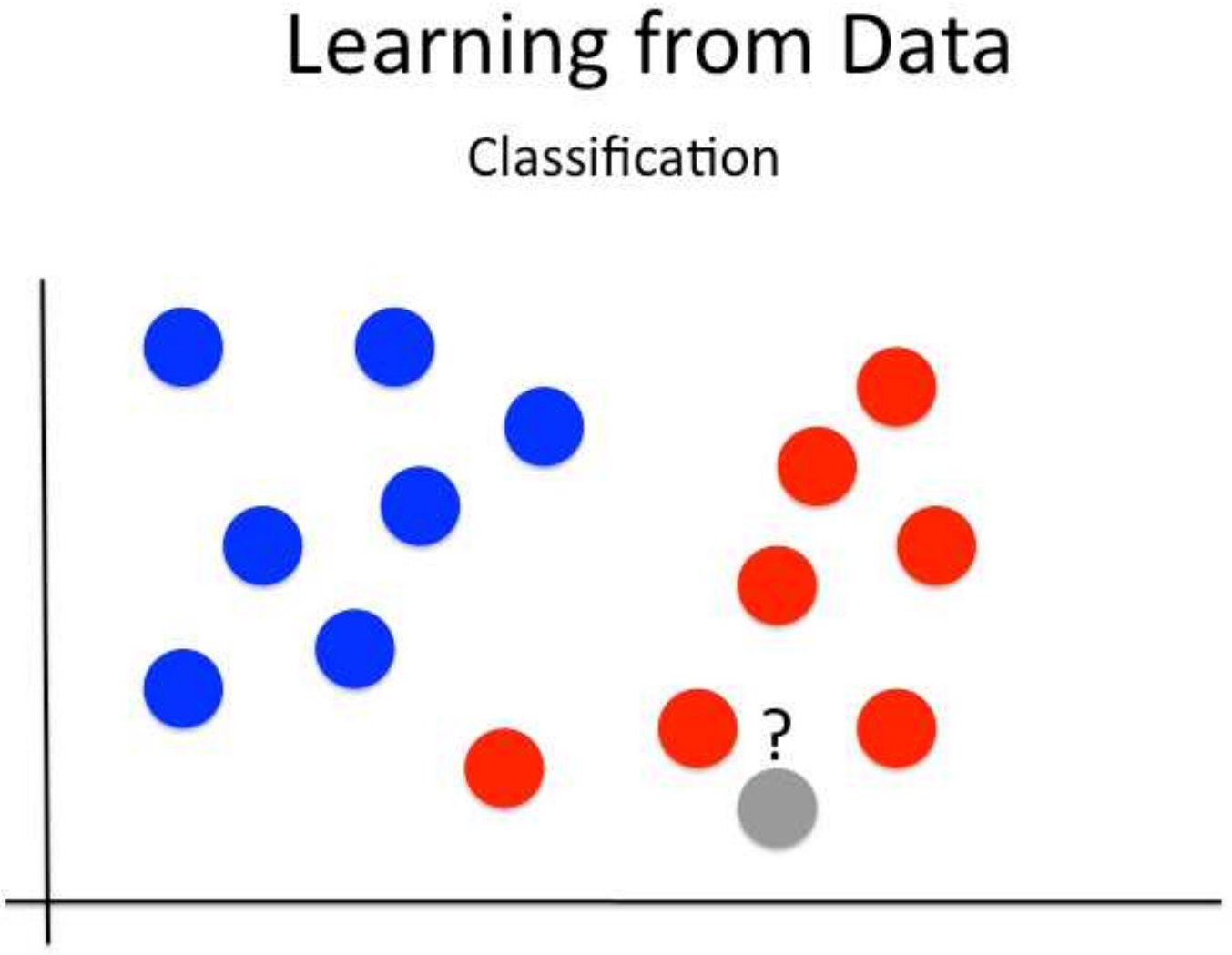
Goal of classification.

Identify a function y, such that $y(\mathbf{x}) = t$.

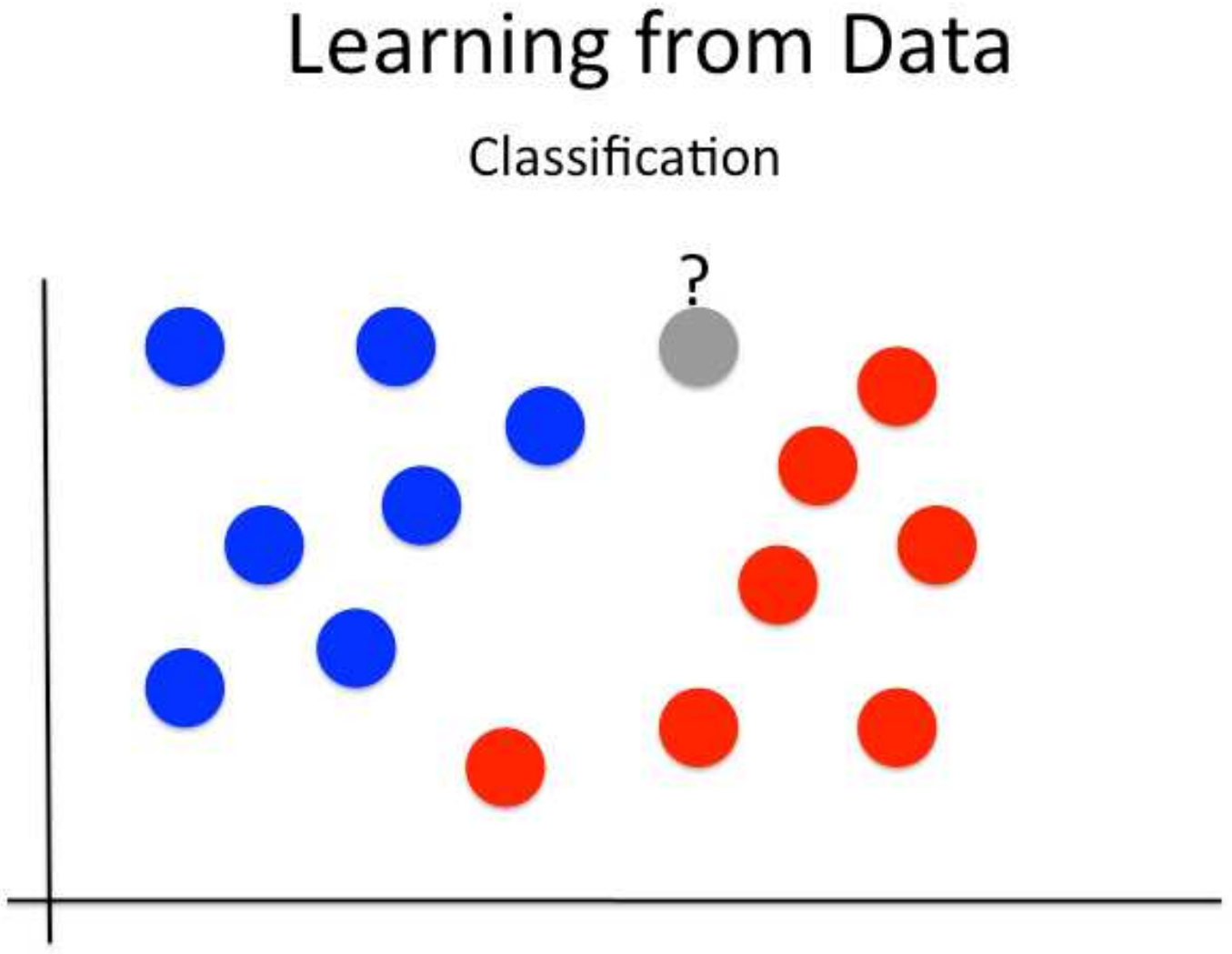
Graphical Example of Classification



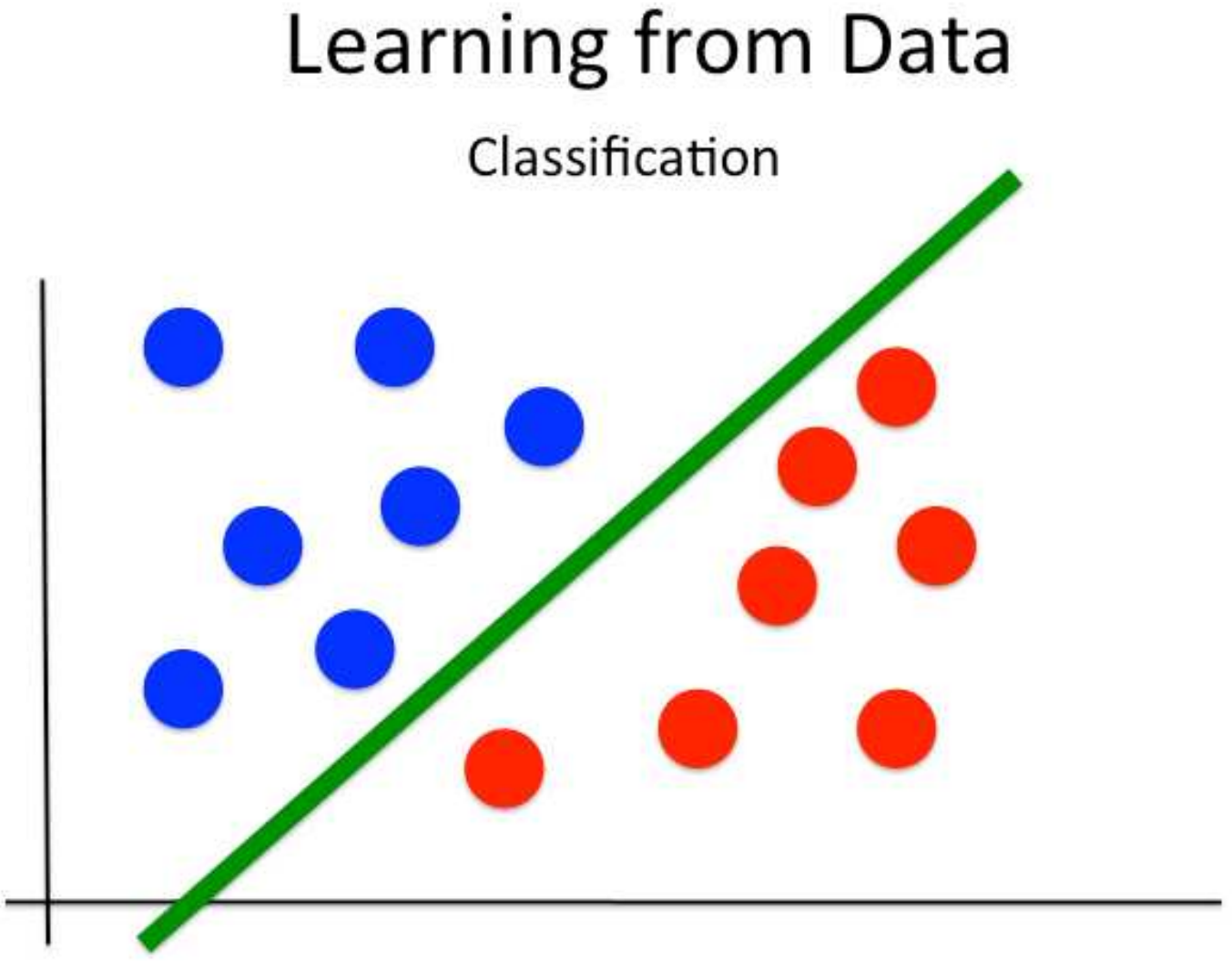
Graphical Example of Classification



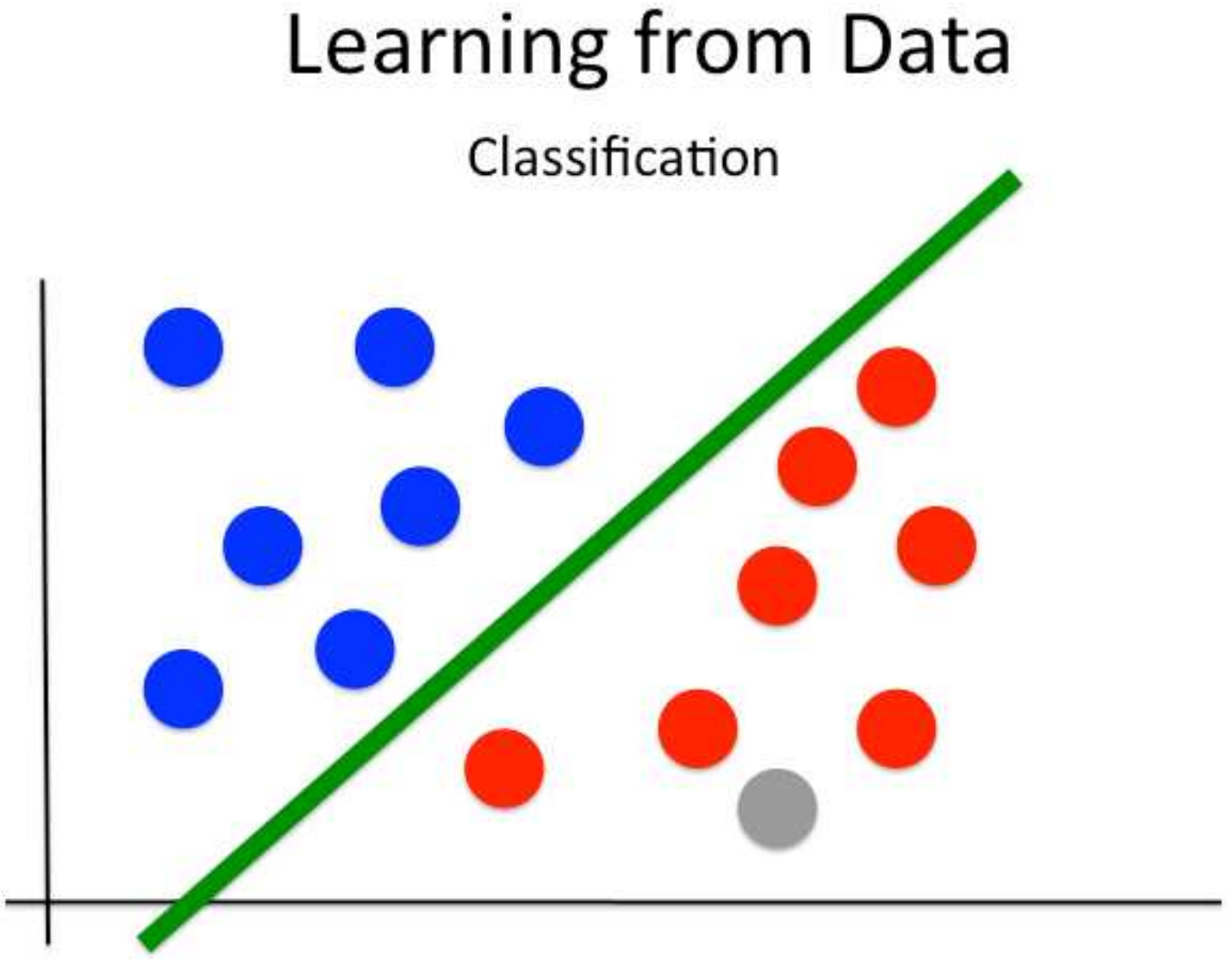
Graphical Example of Classification



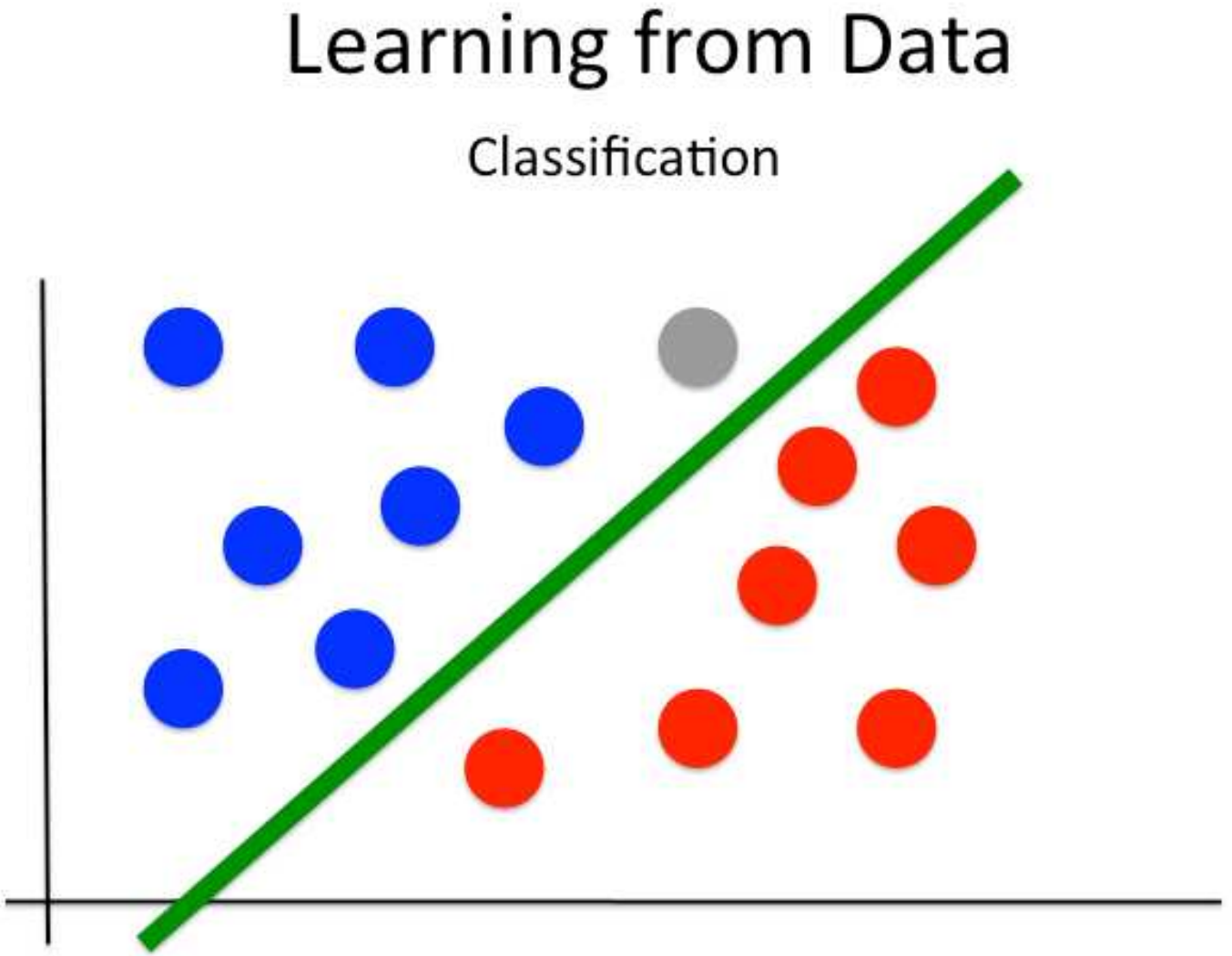
Graphical Example of Classification



Graphical Example of Classification



Graphical Example of Classification



Regression

Regression is another supervised machine learning task.

In classification t was a discrete variable, representing the **class** of the data point, in regression t is a continuous variable.

Goal of regression.

Identify a function y, such that $y(\mathbf{x}) = t$.

Regression

Regression is another supervised machine learning task.

In classification t was a discrete variable, representing the **class** of the data point, in regression t is a continuous variable.

Goal of regression.

Identify a function y, such that $y(\mathbf{x}) = t$.

If the goals of regression and classification are the same, what is the difference?

Regression

Regression is another supervised machine learning task.

In classification t was a discrete variable, representing the **class** of the data point, in regression t is a continuous variable.

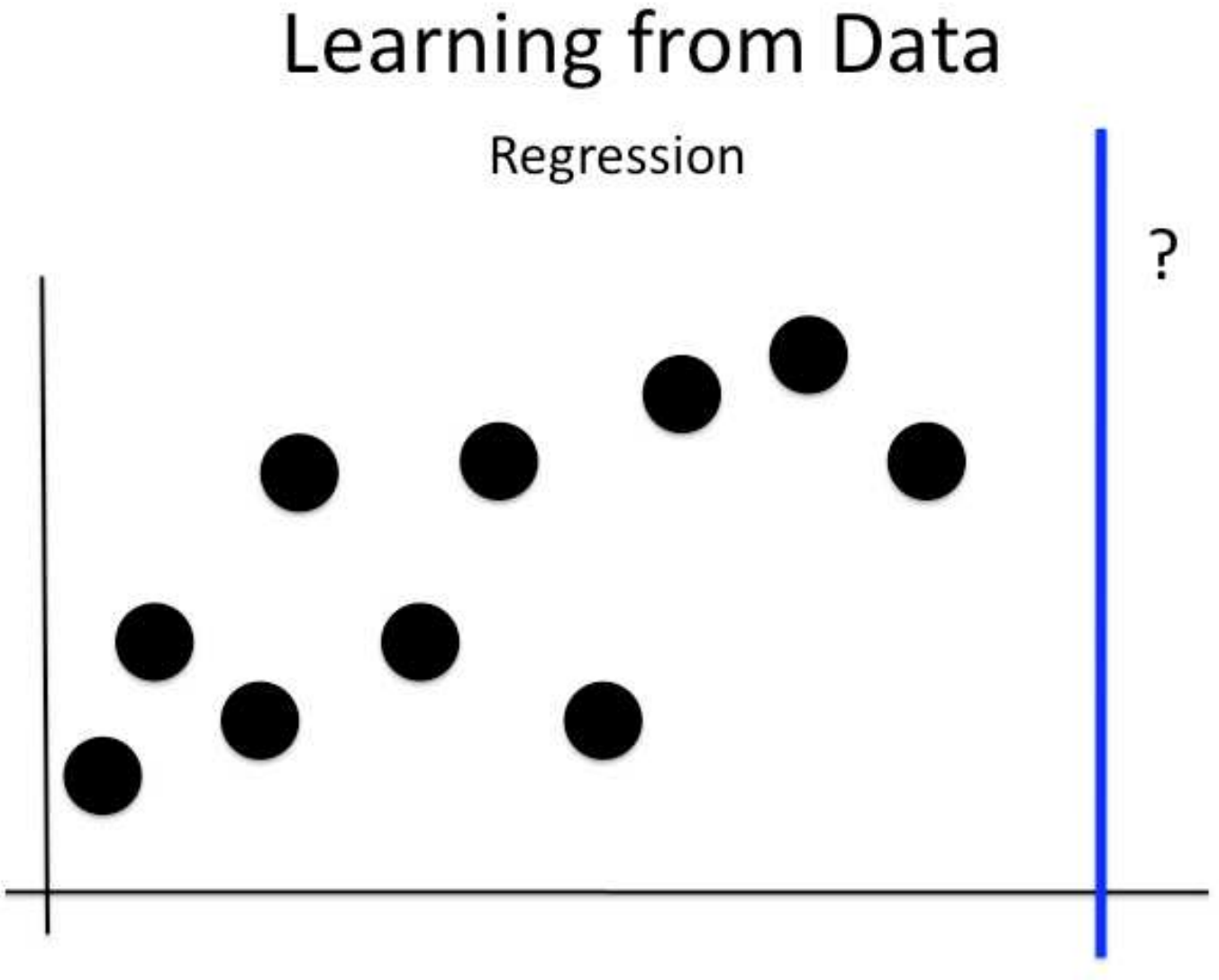
Goal of regression.

Identify a function y, such that $y(\mathbf{x}) = t$.

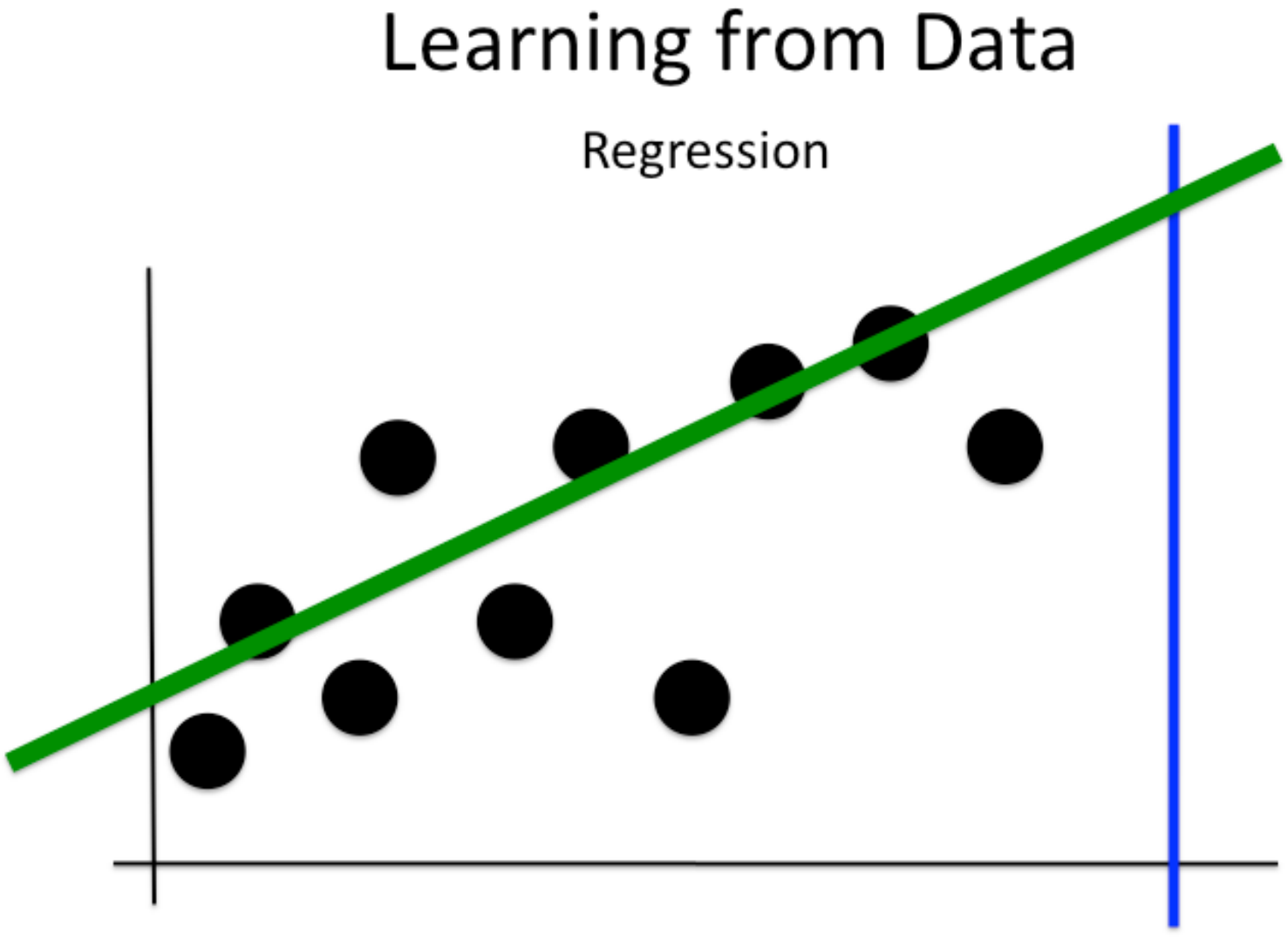
If the goals of regression and classification are the same, what is the difference?

Evaluation.

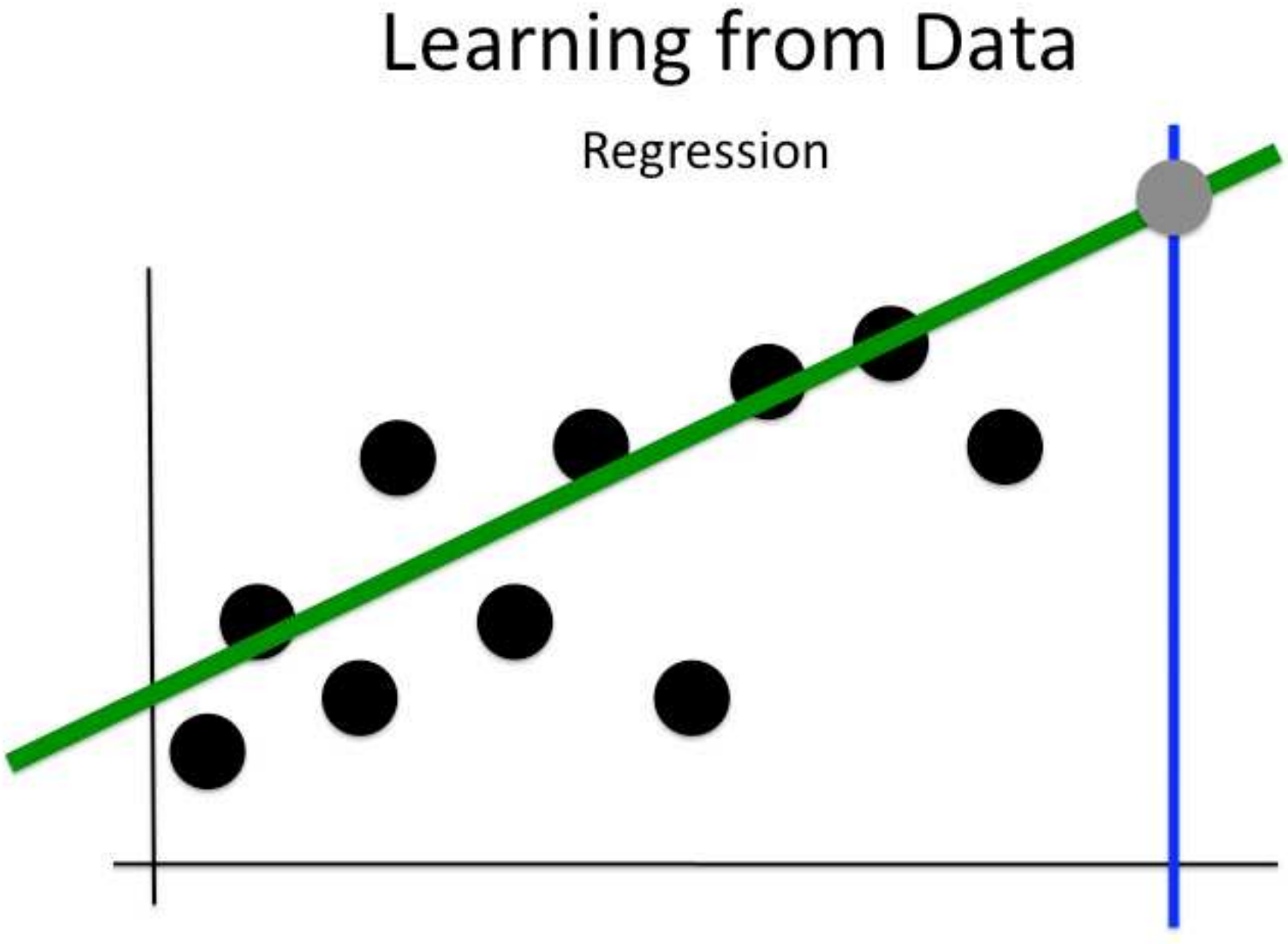
Graphical Example of Regression



Graphical Example of Regression



Graphical Example of Regression



Clustering

Clustering is an **unsupervised** task.

Therefore we have no "target" information to learn.

Rather, the goal is to identify groups of similar data points, that are dissimilar than others.

Technically, identify a **partition** of the data satisfying these two constraints.

1. Points in the same cluster should be **similar**
2. Points in different clusters should be **dissimilar**

Clustering

Clustering is an **unsupervised** task.

Therefore we have no "target" information to learn.

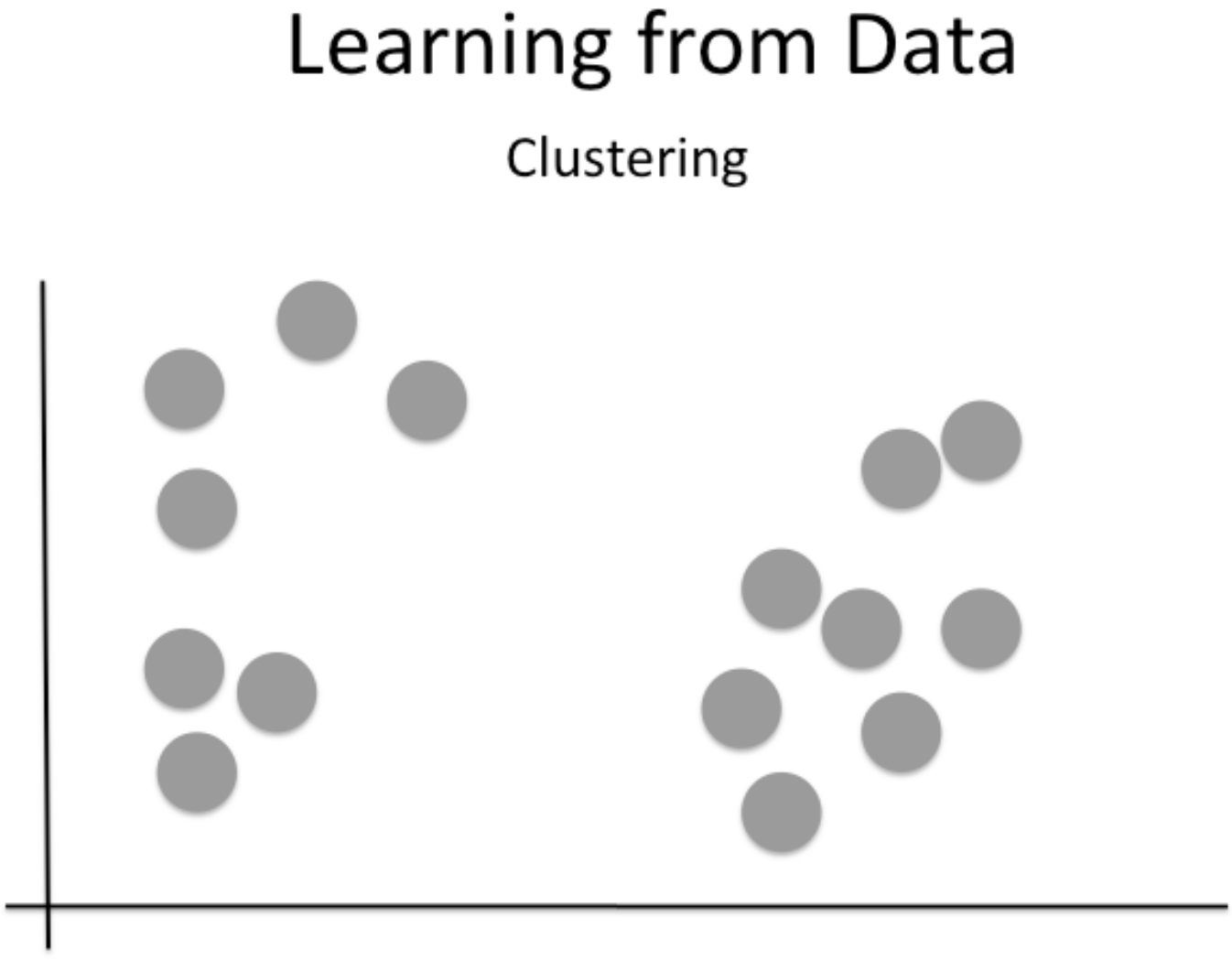
Rather, the goal is to identify groups of similar data points, that are dissimilar than others.

Technically, identify a **partition** of the data satisfying these two constraints.

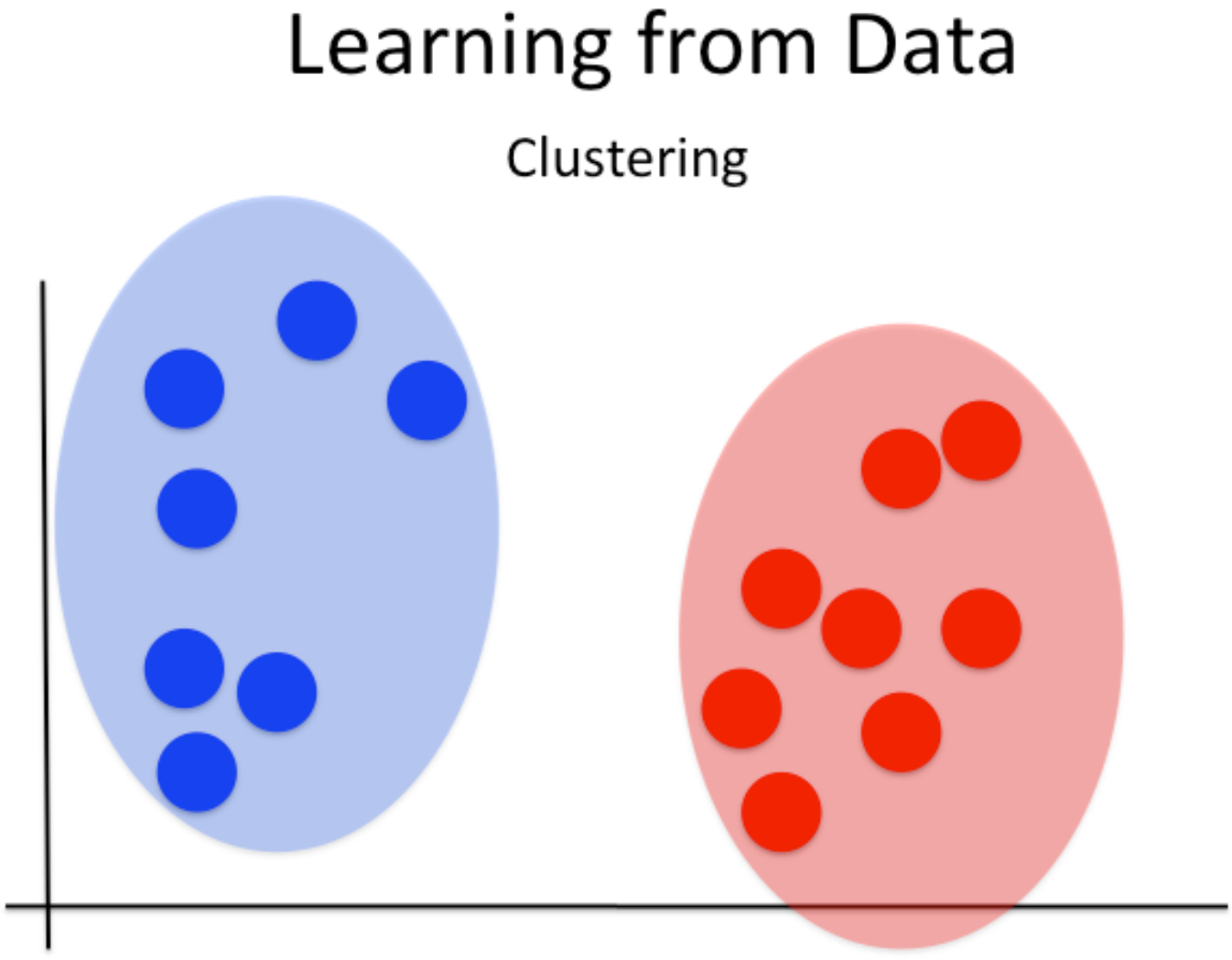
1. Points in the same cluster should be **similar**
2. Points in different clusters should be **dissimilar**

Now the tricky part: Define "Similar".

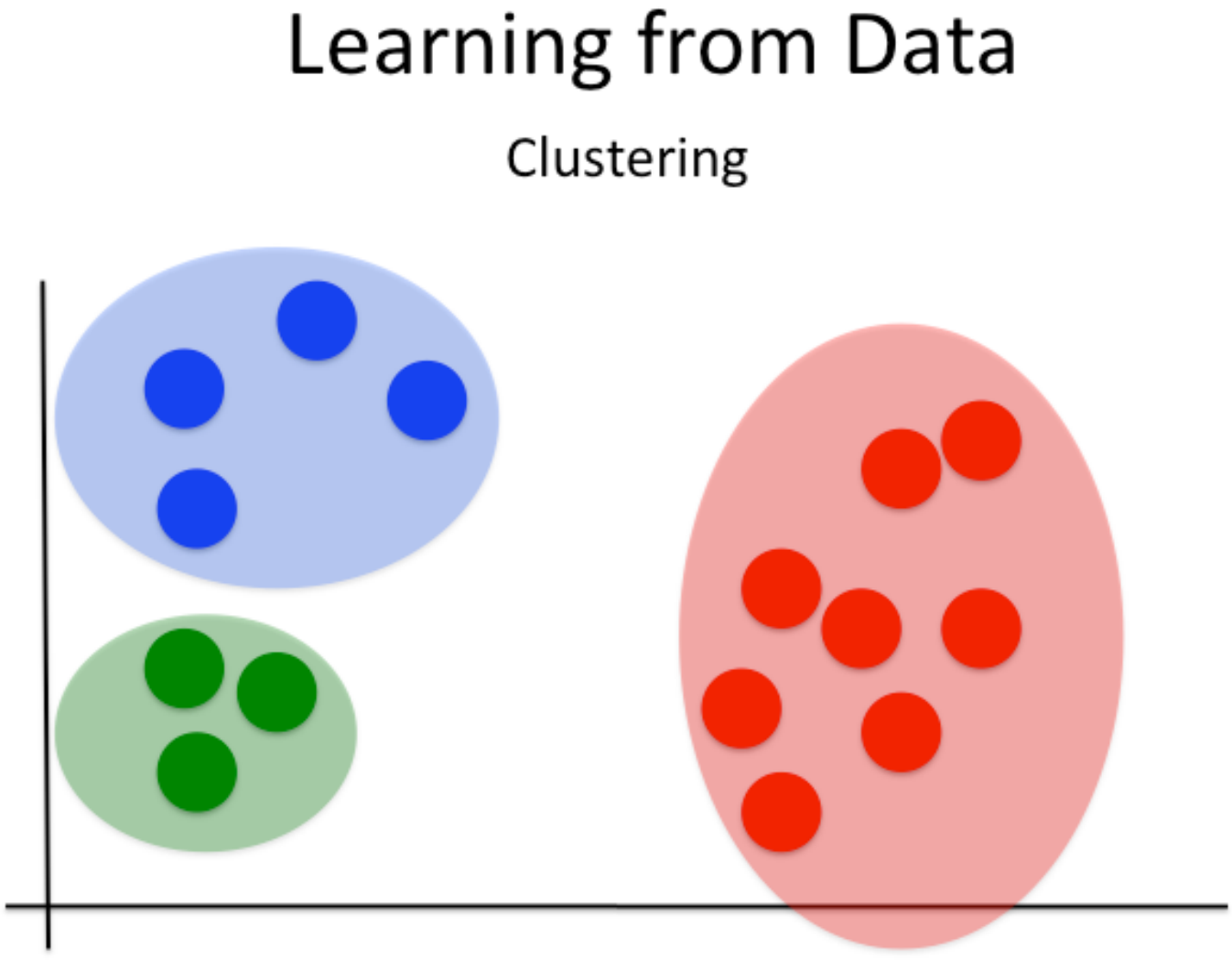
Graphical Example of Clustering



Graphical Example of Clustering



Graphical Example of Clustering



How do we do this?

Mechanisms of Machine Learning.

- Feature Extraction
- Statistical Estimation

Mathematical Underpinnings

What Math will we use?

- Probability and Statistics
- Calculus
- Linear Algebra

Why do we need such complicated math?

How much math?

- A lot.

One common function we will use is the Gaussian Distribution.

$$N(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$

We will be differentiating and integrating over this function.

Why do we need such complicated math?

How much math?

- A lot.

We also look at higher-dimensional Gaussians

$$N(x|\mu,\Sigma) = \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} exp\left\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right\}$$

We will be differentiating and integrating over this function, too.

Policies and Structure

- Course website:
  http://eniac.cs.qc.cuny.edu/andrew/gcml/syllabus.html

Data Data Data

All of the work we will do in this class relies on the availability of data to process.

- UCI: http://archive.ics.uci.edu/ml/
- Netflix Prize: http://archive.ics.uci.edu/ml/datasets/Netflix+Prize
- LDC (Linguistic Data Consortium): http://www.ldc.upenn.edu/

Bye

- Next
  - Probability Review!
    - Frequentists v. Bayesians
    - Bayes Rule