Machine Learning 101

Data as Points in Space

Syllabus

- What is Machine Learning?
- Data as Points in Space
- Basic Relationships (magnitude and angle)
- Objective Functions (what is the good?)
- Two Basic Models
- Introduction to Decision Trees
- Overview of VenueSearch

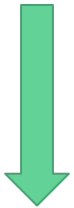
What is Machine Learning?

- Means of using the **observed** to make inferences about the **unobserved**.
- Means of enabling computers to **program themselves**.
- Used at **Foursquare** to:
  - Determine venue visits from phone signals
  - Aggregate and reconcile venue data from across the web
  - Identify tip sentiment (positive or negative)
  - Predict user age/gender from past visits
  - And more!

Key Concepts (to look out for)

- Data
  - Features
  - Labels
- Model
- Objective function
- Model fitting
  - Training data
  - Testing data
  - Overfitting

Representation



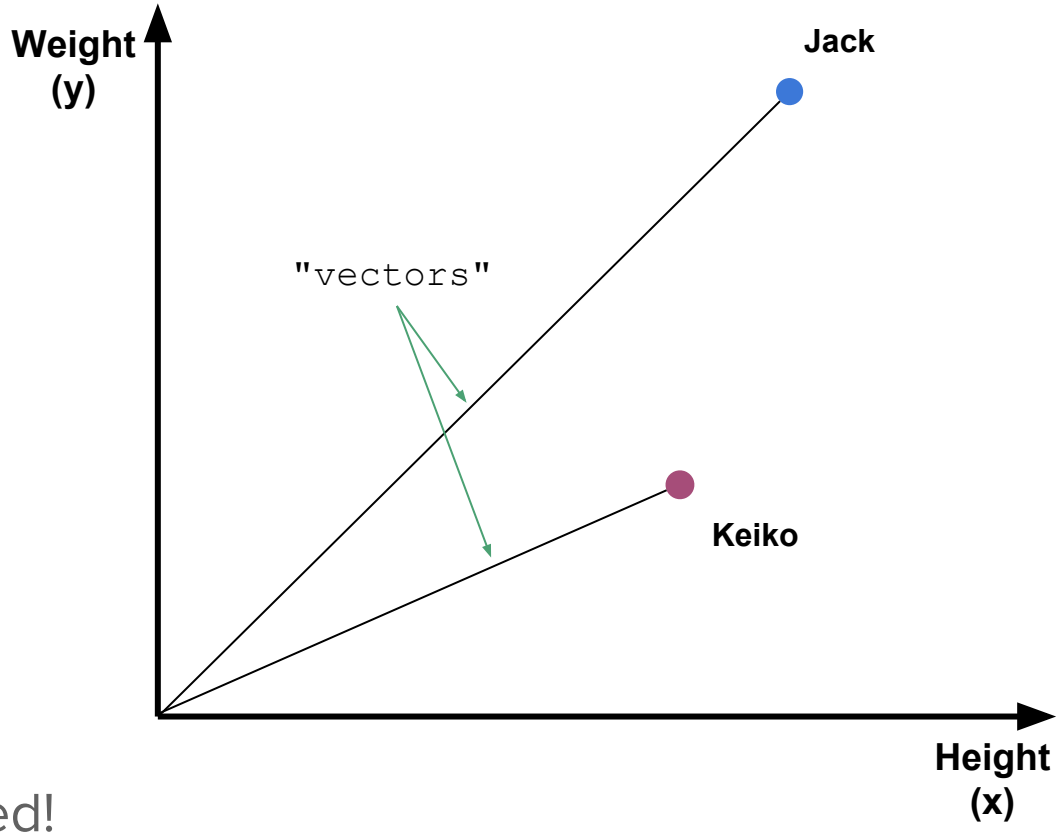
Analysis

Big idea

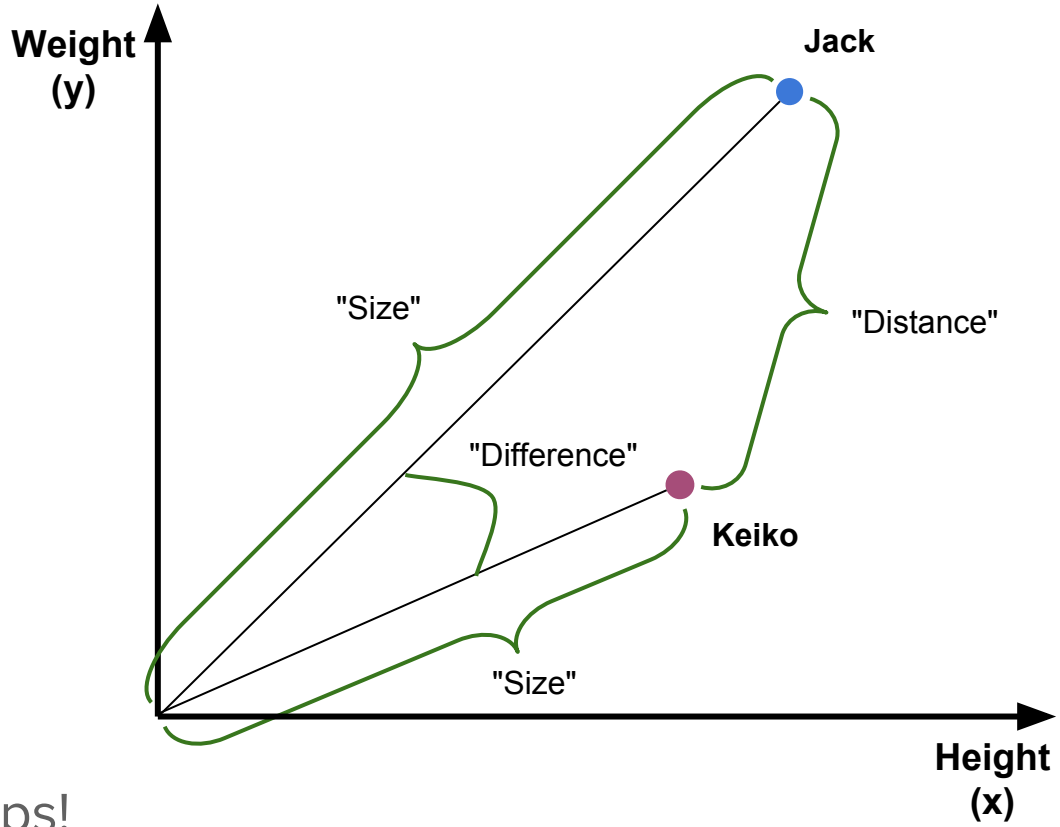
`metadata` `metadata` `label` `feature` `feature`

ID	Name	Gender	Height (in)	Weight (lb)
1	Jack	M	70	180
2	Keiko	F	62	130
...	...	...	...	...

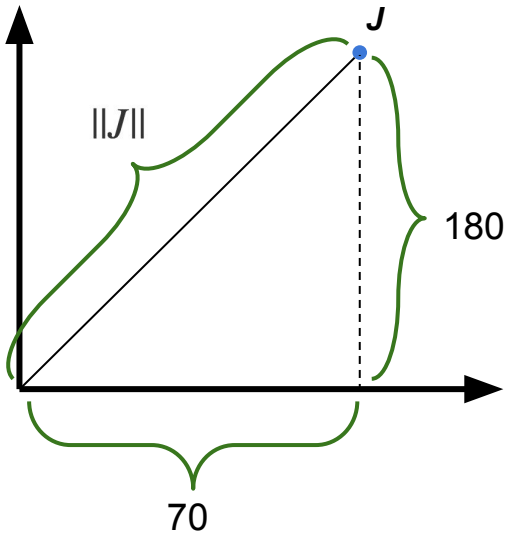
Data!



Transformed!



Relationships!



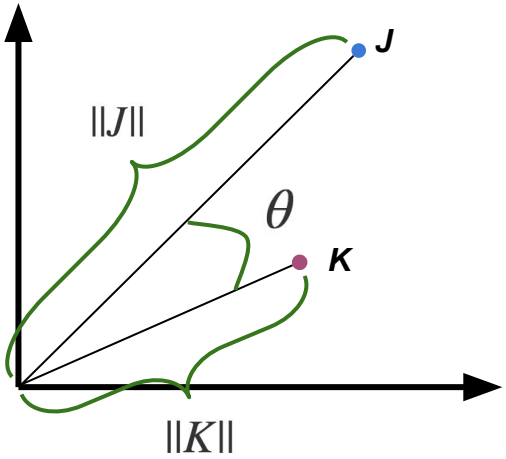
$$J = [70, 180]$$

$$J \in \mathbb{R}^2$$

$$\|J\| = \sqrt{\sum_{i=1}^{d} J_i^2}$$

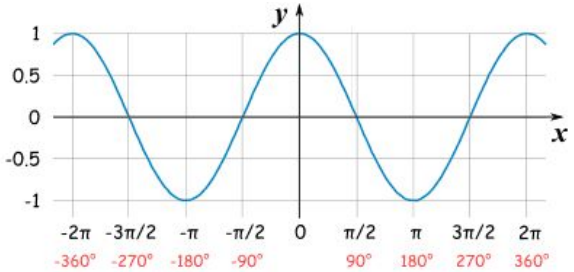
this is just the pythagorean function ;)

Meet the Vector



$$J^T K \triangleq \sum_{i=1}^{d} J_i \cdot K_i$$

$$= ||J|| \cdot ||K|| \cdot cos(\theta)$$



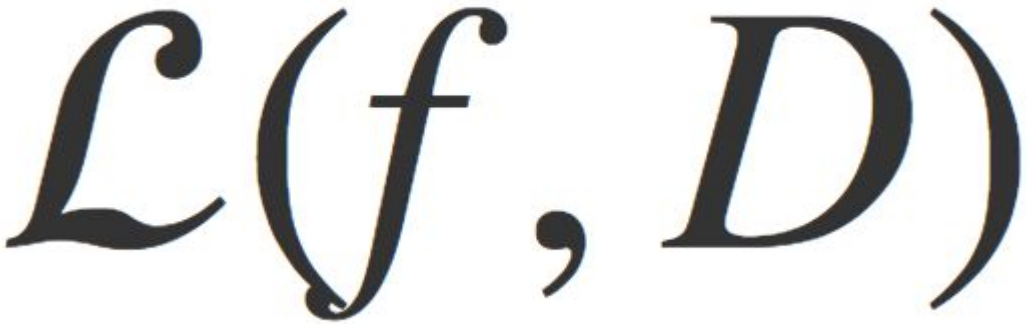
Meet the Dot Product

So.

- Data are a set of points (vectors) in (some dimensional) space.
    - We call a vector of vectors a matrix.
- These points have (numeric) properties, mostly from trigonometry.
- Machine learning "models" are functions of these properties:

To know the good is to desire the good.
- Socrates

What is "the good" ?



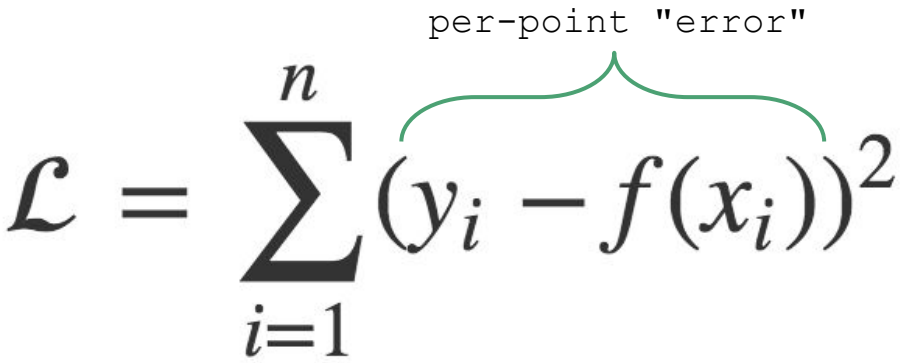
The "Objective Function" (aka "Loss Function")

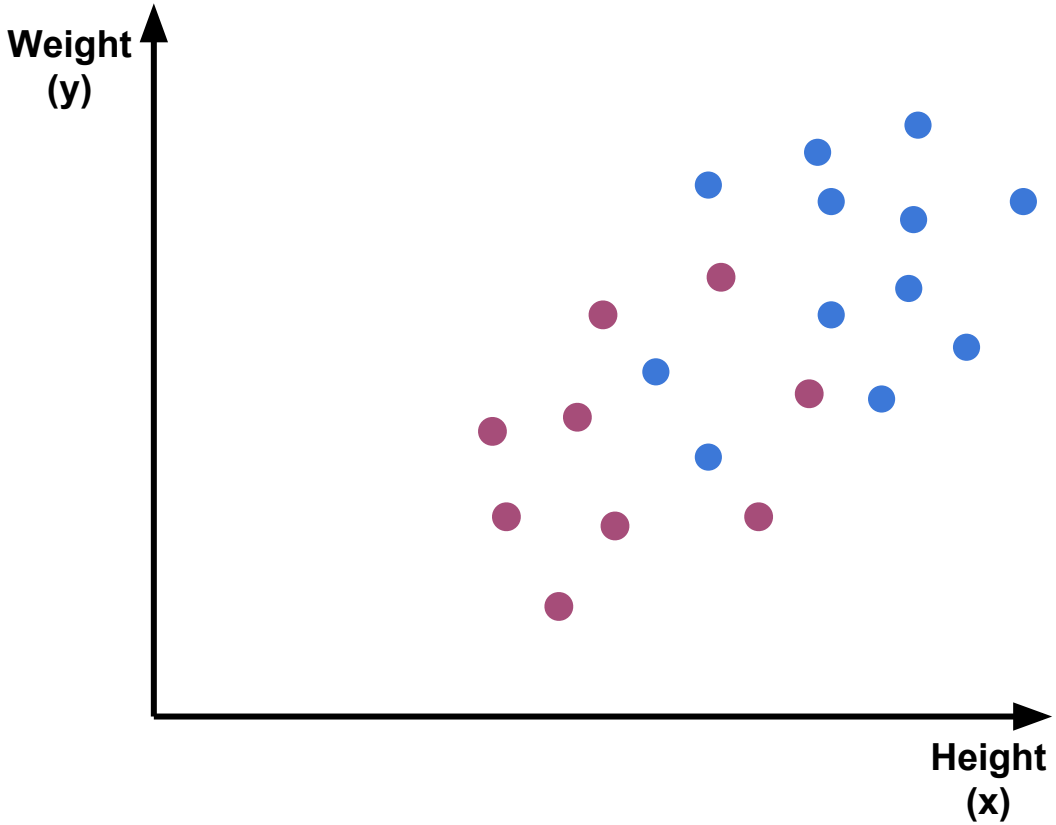
Idea:

- The objective function tells us how "good" the model is.
- These functions come in many flavors. Two common ones are:
    - Measurements of **error** (which we want to **minimize**).
    - Measurements of **likelihood** (which we want to **maximize**).
- We can use various techniques to **optimize** the model w.r.t. the objective.
- We call this process "**fitting**" a model (to data).

Example: Least-Squares Regression

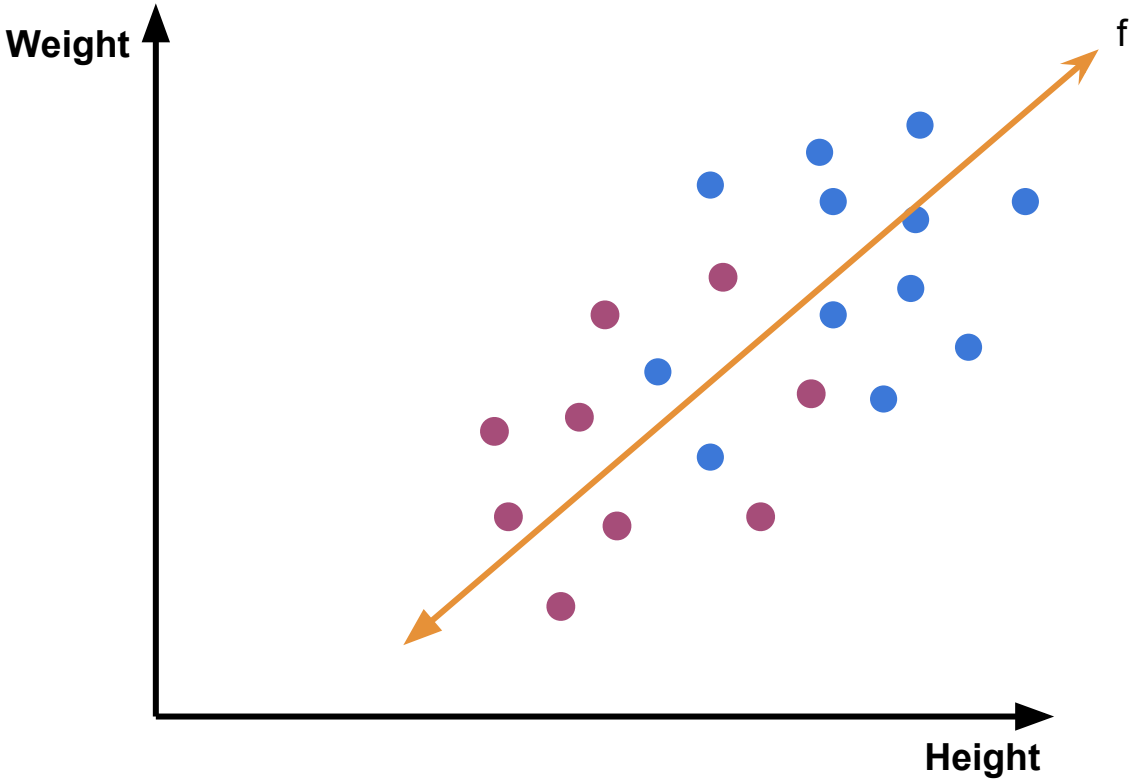
- **Idea**: features of a data point are related in some way.
- Example: predict MPG for a car using year, weight, engine size, etc.
- Objective function:





Moar Data!

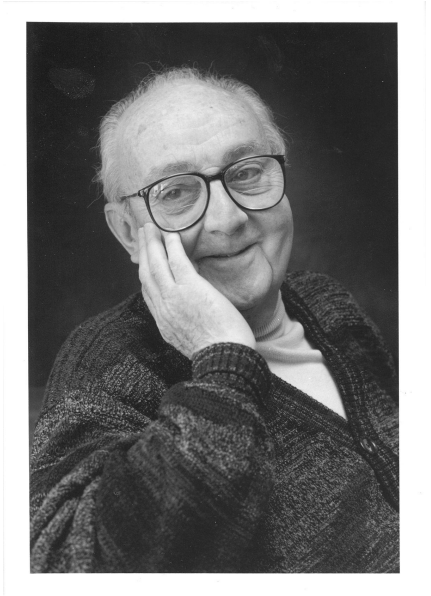
Optimize!



Bam!

Note:

- Obviously, the real relationship between height and weight is very complex.
- A model is a simplification of the real world.
- ***"All models are wrong, but some are useful"***
  - George Box

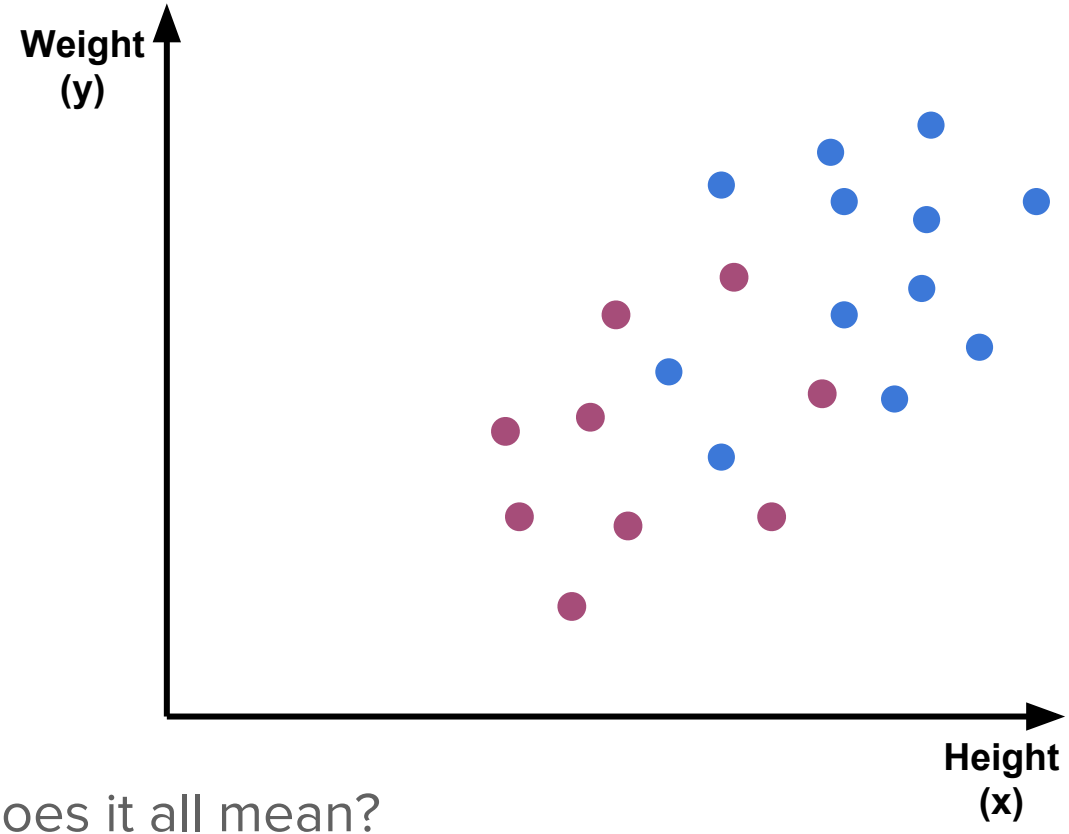


Example: Logistic Regression

- **Idea**: the features of a data point can tell us about the **class** of thing it is.
- Used at **Foursquare** for sentiment analysis (positive or negative sentiment?).
- Objective function:

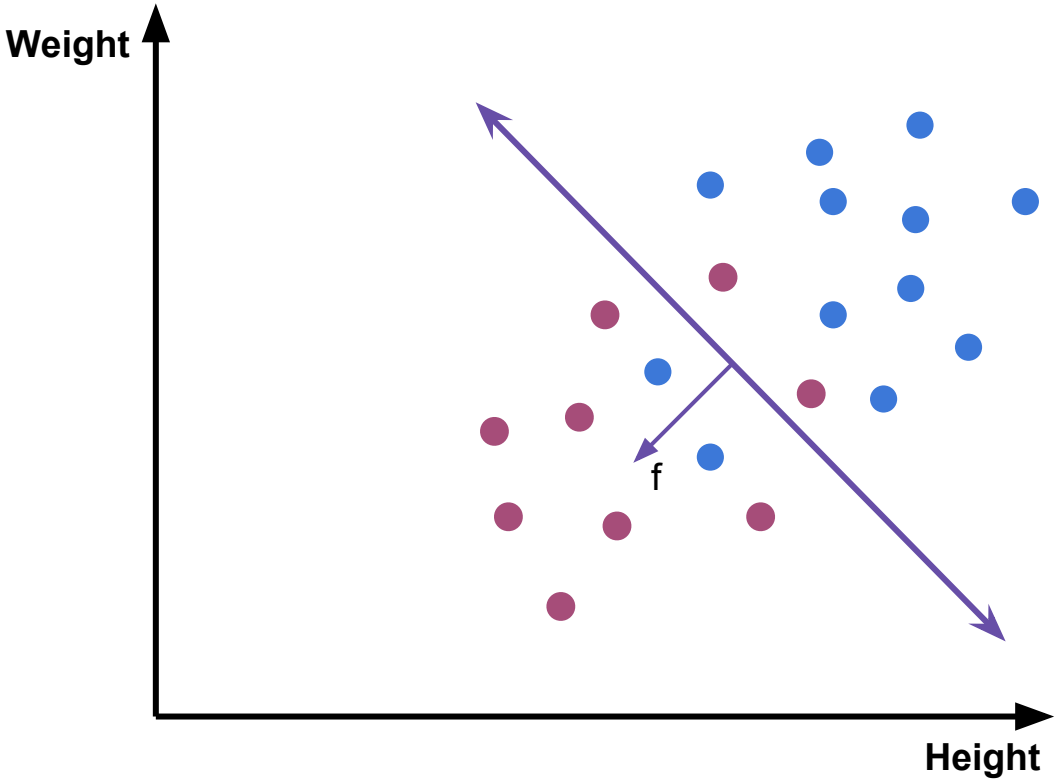
probability of right class

$$\mathcal{L} = \sum_{i=1}^{n} log(f(x_i, y_i) \cdot c_i)$$



But what does it all mean?

Optimize!



Wham!

Summary so far:

- Data can be thought of as points (vectors) in (some dimensional) space.
- A machine learning system consists of the following building blocks:
  - A model (a simplified view of some real-world relationship)
  - An objective function (a way to measure how good a model is, relative to data)
  - An algorithm for updating the model with regards to the objective.
- Neural networks, etc are just fancier versions of this.
- No magic anywhere! Except the dot product.