Neuroscience Using Machine Learning Concepts

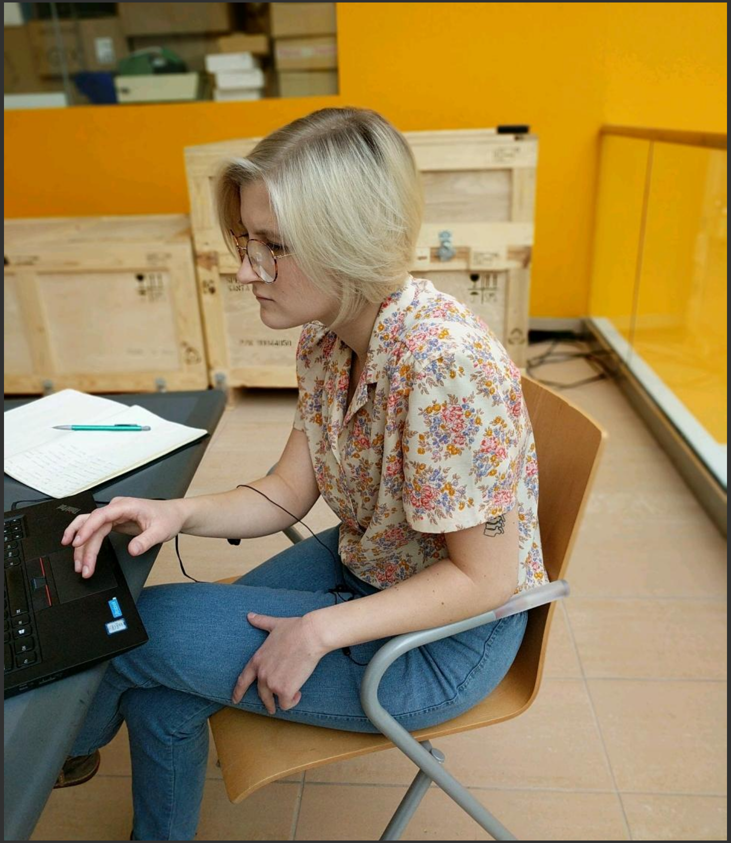
2020 Quantitative Methods Workshop at MIT

Taylor Baum

Taylor Baum

tbaum@mit.edu
taylor.e.baum@gmail.com

https://www.linkedin.com/in/taylor-baum-b1a4aa105/



What is Machine Learning?

An algorithm is similar to a recipe.

```
1. Preheat oven to 375 degrees F.
2. Line a baking pan with parchment
   paper and set aside.
3. In a separate bowl mix flour,
   baking soda, salt, baking powder.
   Set aside.
4. Cream together butter and sugars
   until combined.
5. ...
```

More formally, an algorithm is a process or set of rules with which a task is completed or with which an input is manipulated to produce an output.

```
if x
   do y;
if else z
   do a;
else
   break;
```

For example, we use algorithms for tasks like sorting.

```
Input: Unsorted Numbers
```

$\Big\downarrow$ `Algorithm?`

```
Output: Sorted Numbers
```

What if there is a task where the algorithm isn't obvious?

```
Input: Emails
   ↓ Algorithm?
Output: Indication of Spam (Yes/No)
```

What can we do?

Input: Emails

Algorithm?

Output: Indication of Spam (Yes/No)

Well, we often have a large amount of data that we can use to help inform how we produce appropriate outputs from these inputs.

DATA

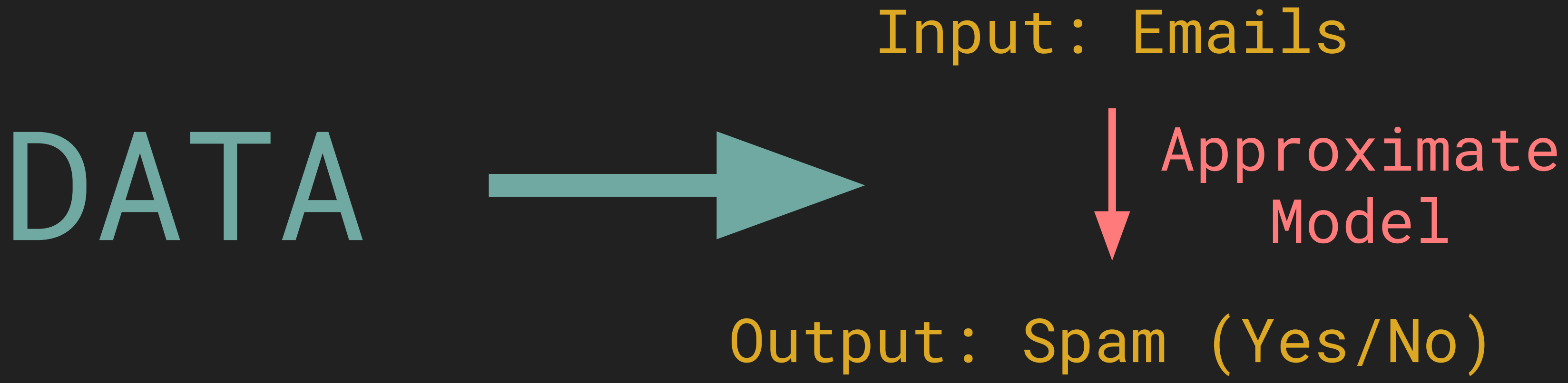
Machine Learning provides automated methods of data analysis that can then be used further.

Input: Emails

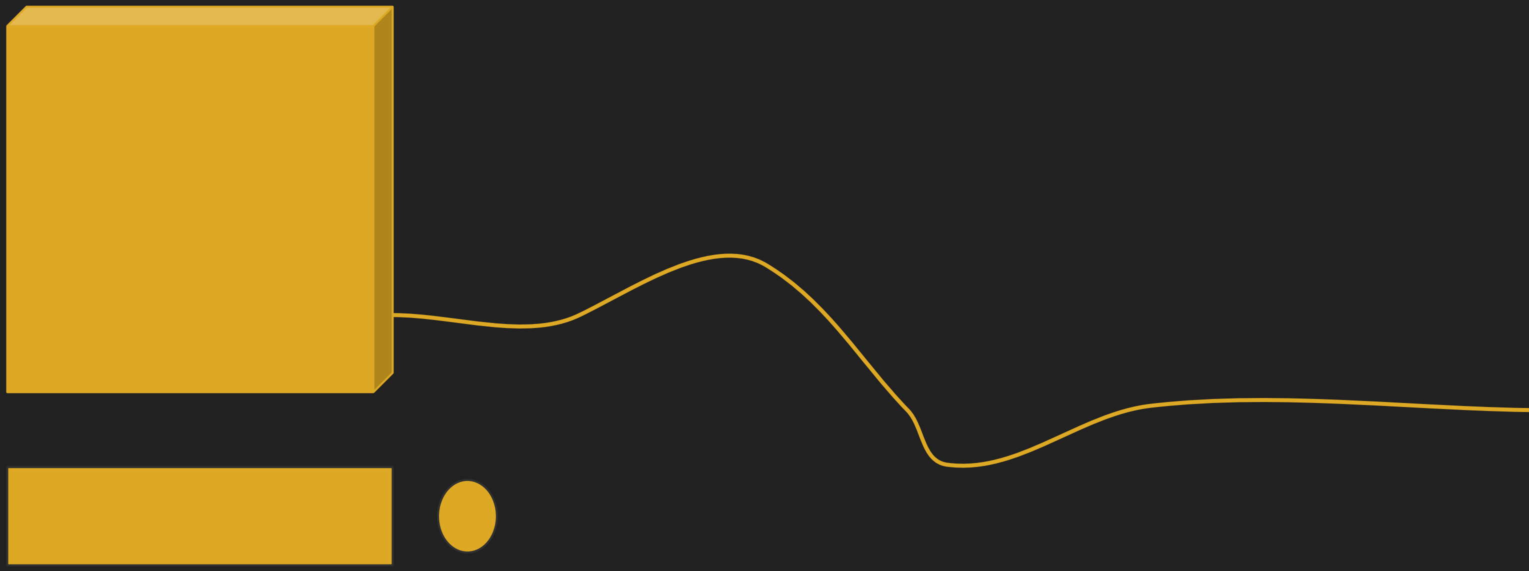
Algorithm?

Output: Spam (Yes/No)

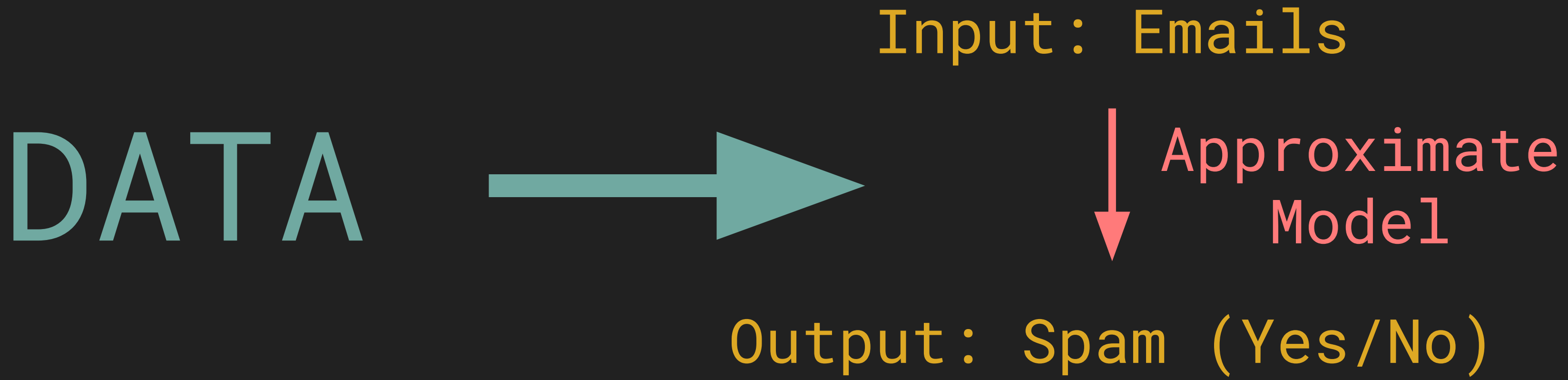
Machine Learning provides automated methods of data analysis that can then be used further.



Machine Learning is a set of methods that can automatically detect patterns in data, and then use the uncovered patterns to predict future data, or to perform other kinds of decision making under uncertainty.

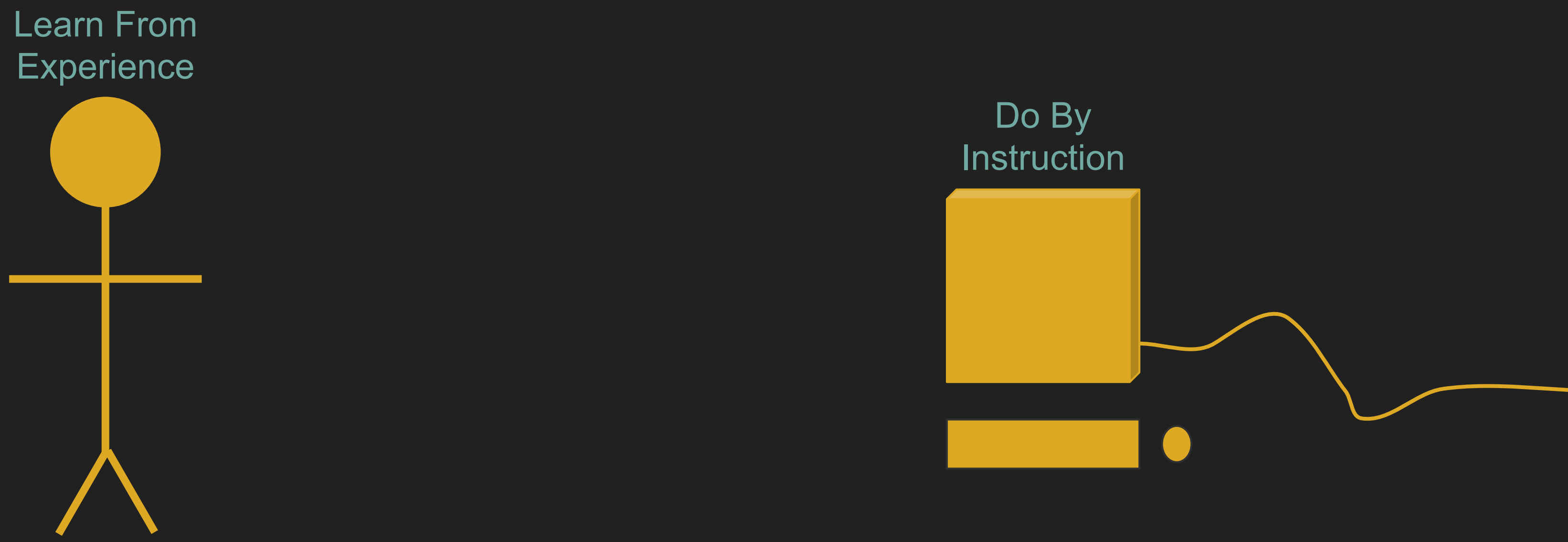


So, we want to learn the patterns or approximate models from the data, and then use that approximate model to make predictions on new data.

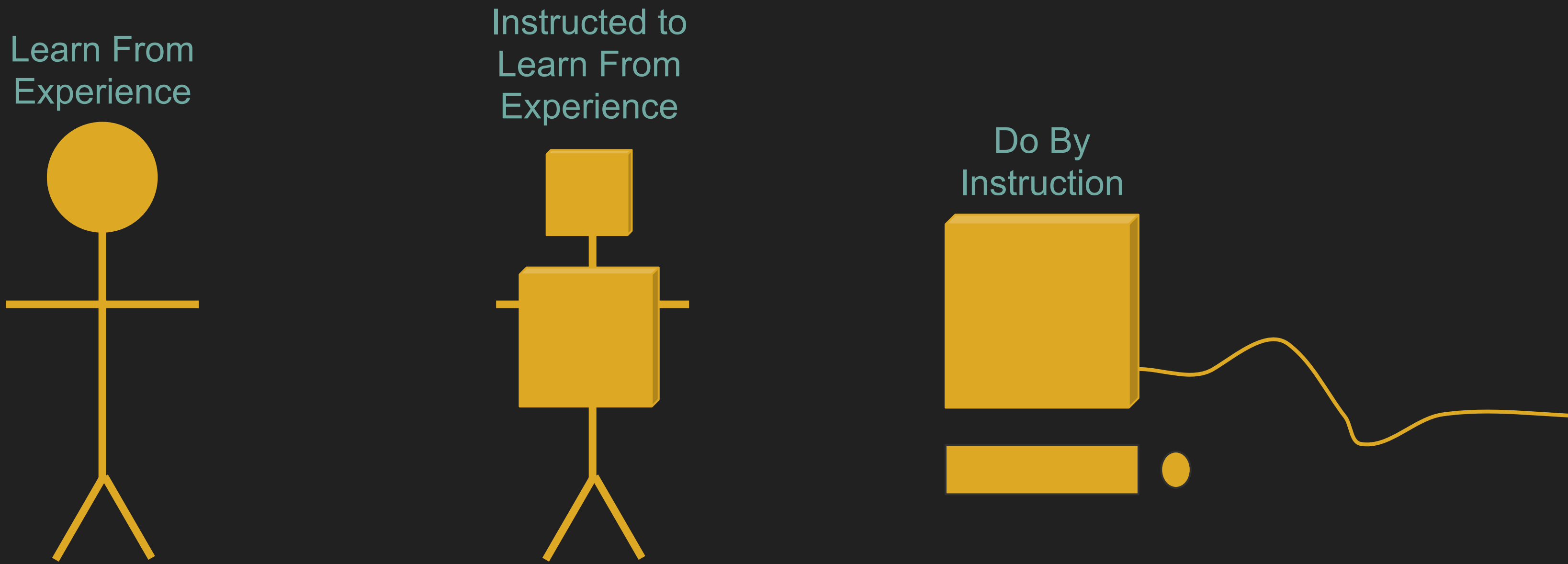


Where did the idea for Machine Learning come from?

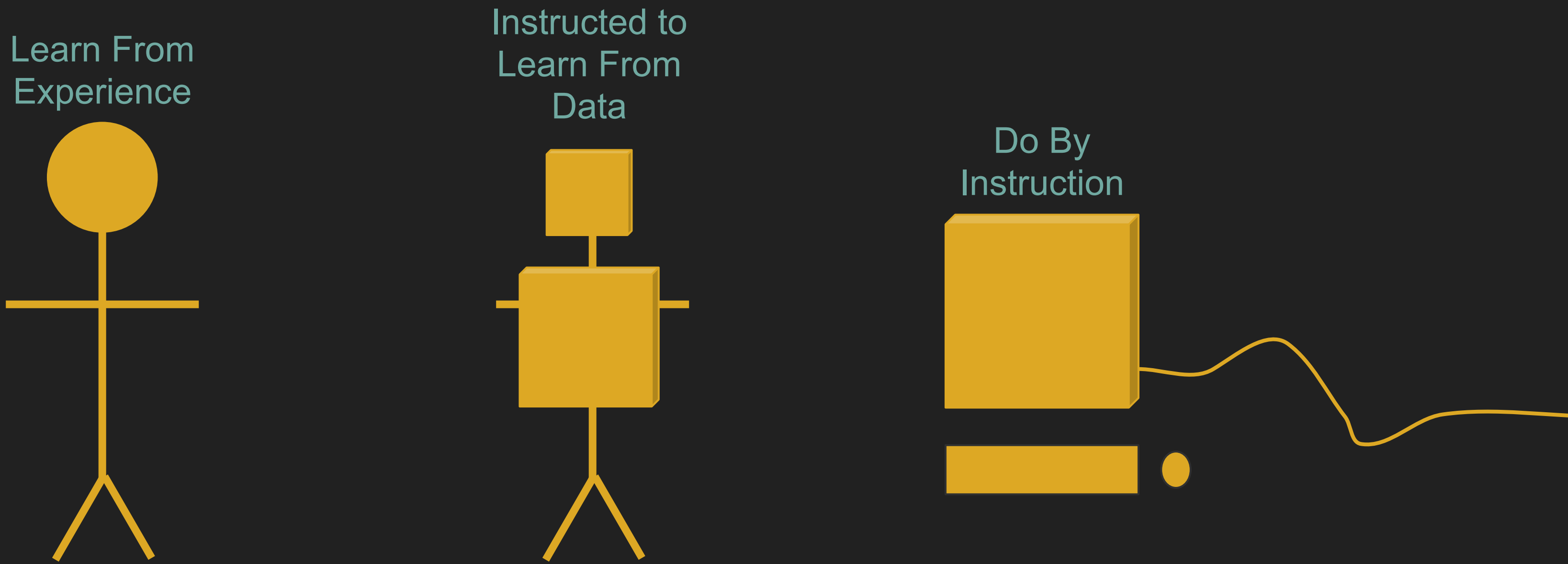
Traditionally, humans learn how to do certain tasks from experience in the real-world while computers must be instructed to do things.



Machine learning is a field which seeks to enable computers and machines to learn from experience.



The experience that computers have access to is known as data.



Thus far, there are three different ways that a machine can learn. We will focus on 1, and 2.

1. Supervised Learning

2. Unsupervised Learning

3. Reinforcement Learning

Supervised learning uses labeled data as it’s feature set, or the data we initially have to learn patterns from.

1.

Supervised Learning

The labeled data is a set of input-output pairs.

1.

Supervised Learning

For example...

$$X \longrightarrow Y$$

For example...

For example...

→ `evil`

For example...

→ cat

In other words, you have input variables (X) and their corresponding output variables (Y) and you want to learn the mapping between these variables.

1.

Supervised Learning

Unsupervised learning uses unlabeled data as it’s feature set, or the data we initially have to learn patterns from.

2.

Unsupervised Learning

In unsupervised learning, we only have a set of inputs, and the goal is to discover interesting patterns in the data.

2.

Unsupervised Learning

We will look at examples of both of these learning techniques!

1\. Supervised Learning

2\. Unsupervised Learning

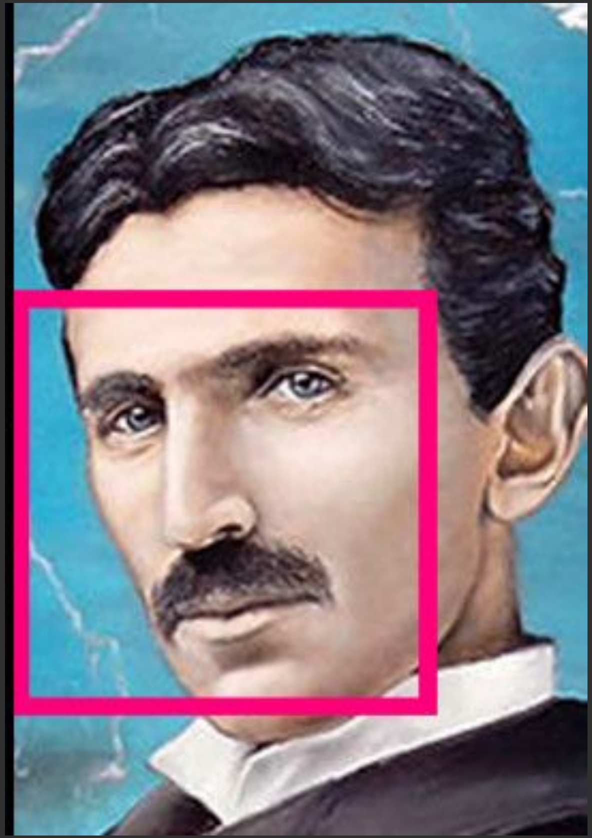
What are some achievements which have arisen from Machine Learning?

What are some achievements which have arisen from Machine Learning?

Alpha Go

What are some achievements which have arisen from Machine Learning?

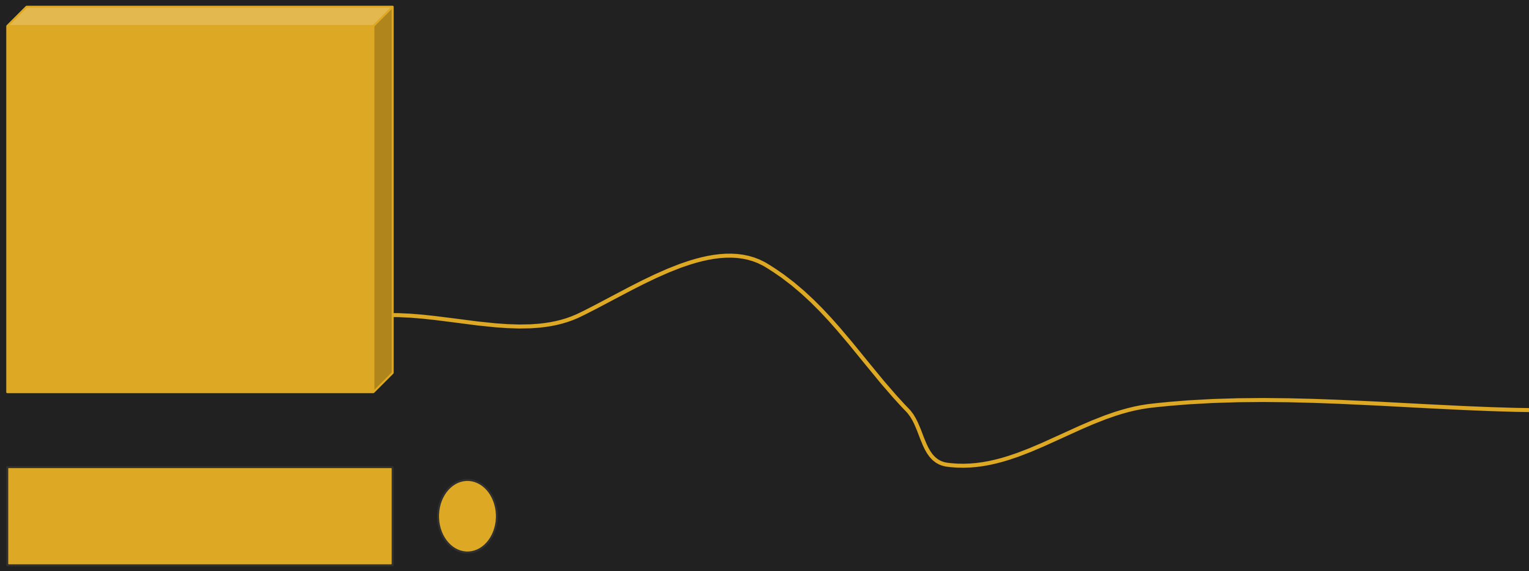
Face Detection



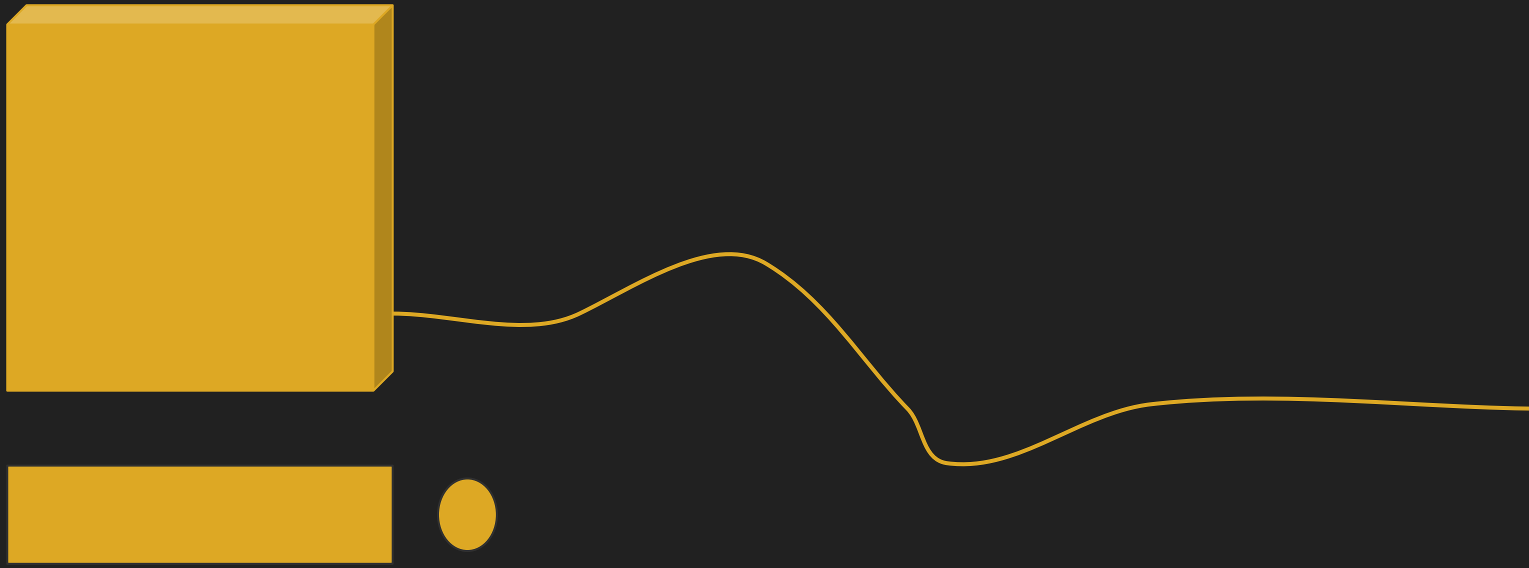
What are some achievements which have arisen from Machine Learning?

Object Detection

In the past three examples, we saw that these tasks are complex, like a game with many rules.



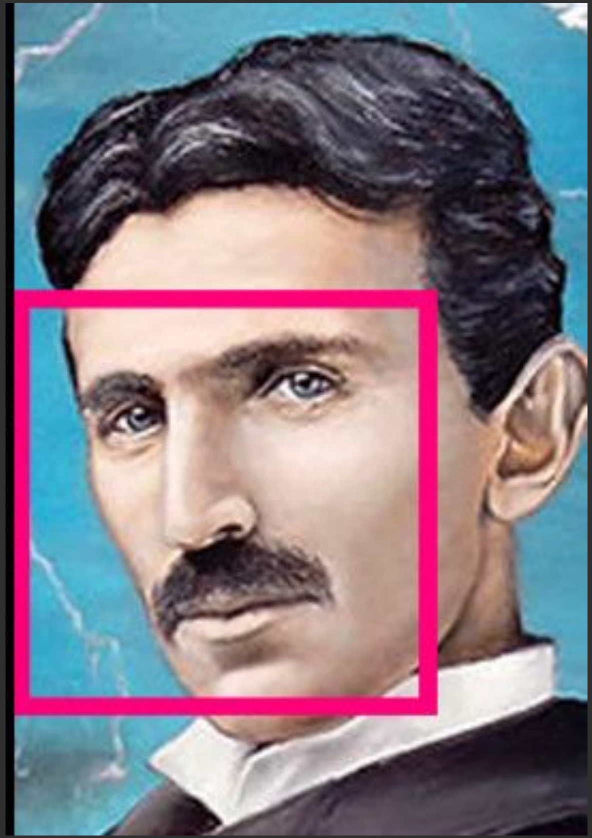
Before Machine Learning approaches, progress towards achieving these tasks was minimal.



What data could have been used for the previous examples?

What data could have been used for the previous examples?

Face Detection



What data could have been used for the previous examples?

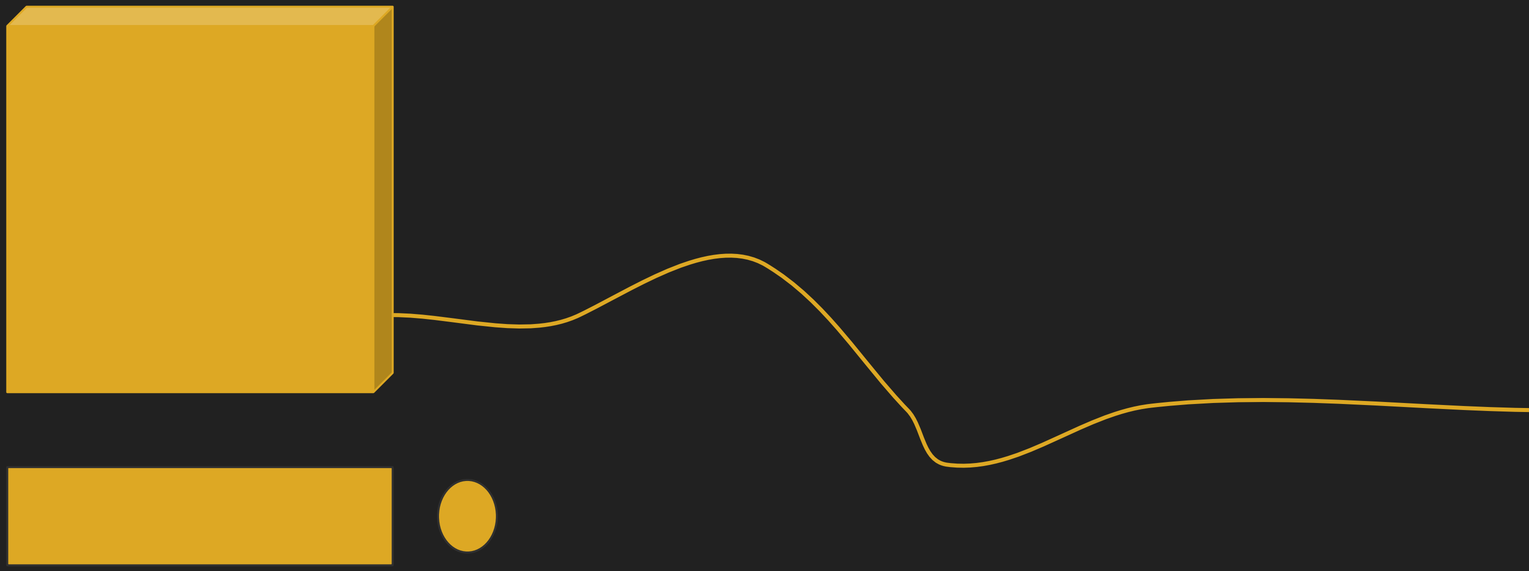
Alpha Go

What data could have been used for the previous examples?

Object Detection

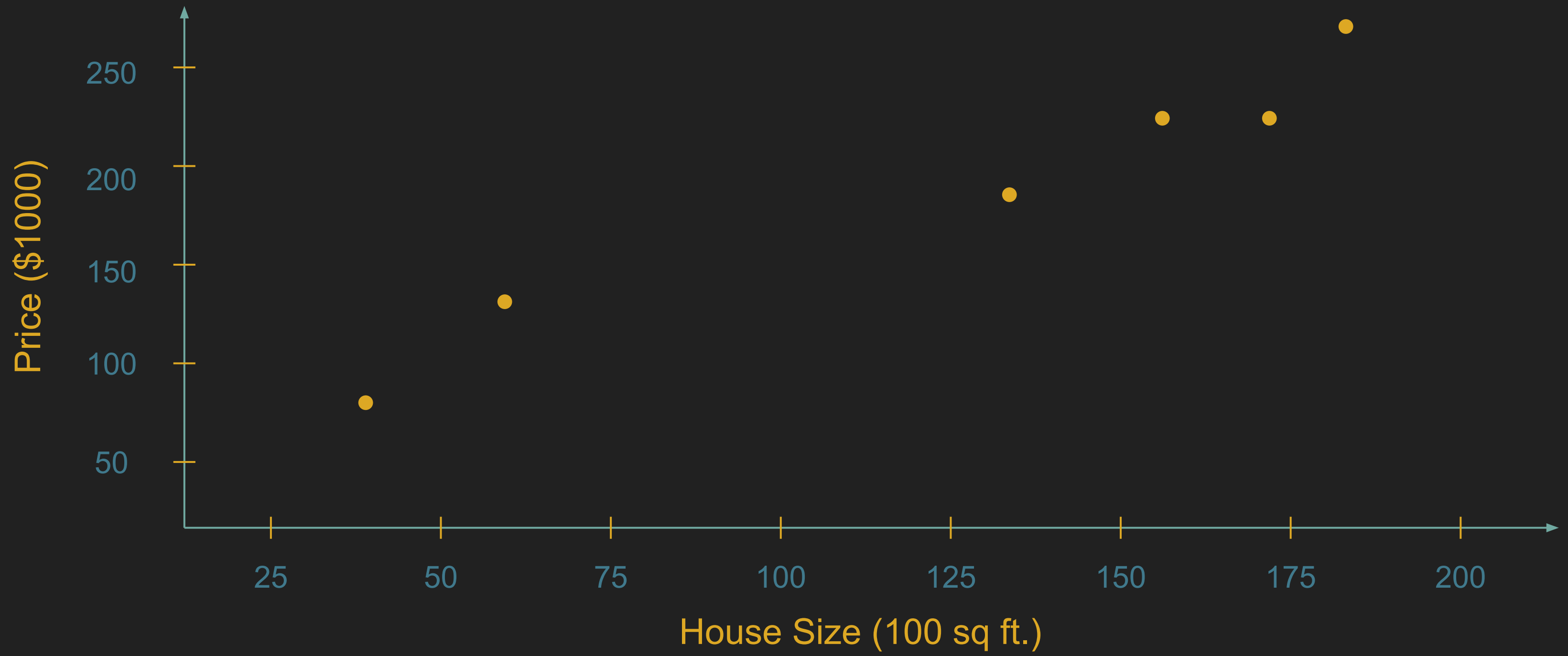
Machine learning is a field which seeks to enable computers and machines to learn from experience or data.

Machine Learning is a set of methods that can automatically detect patterns in data, and then use the uncovered patterns to predict future data, or to perform other kinds of decision making under uncertainty.

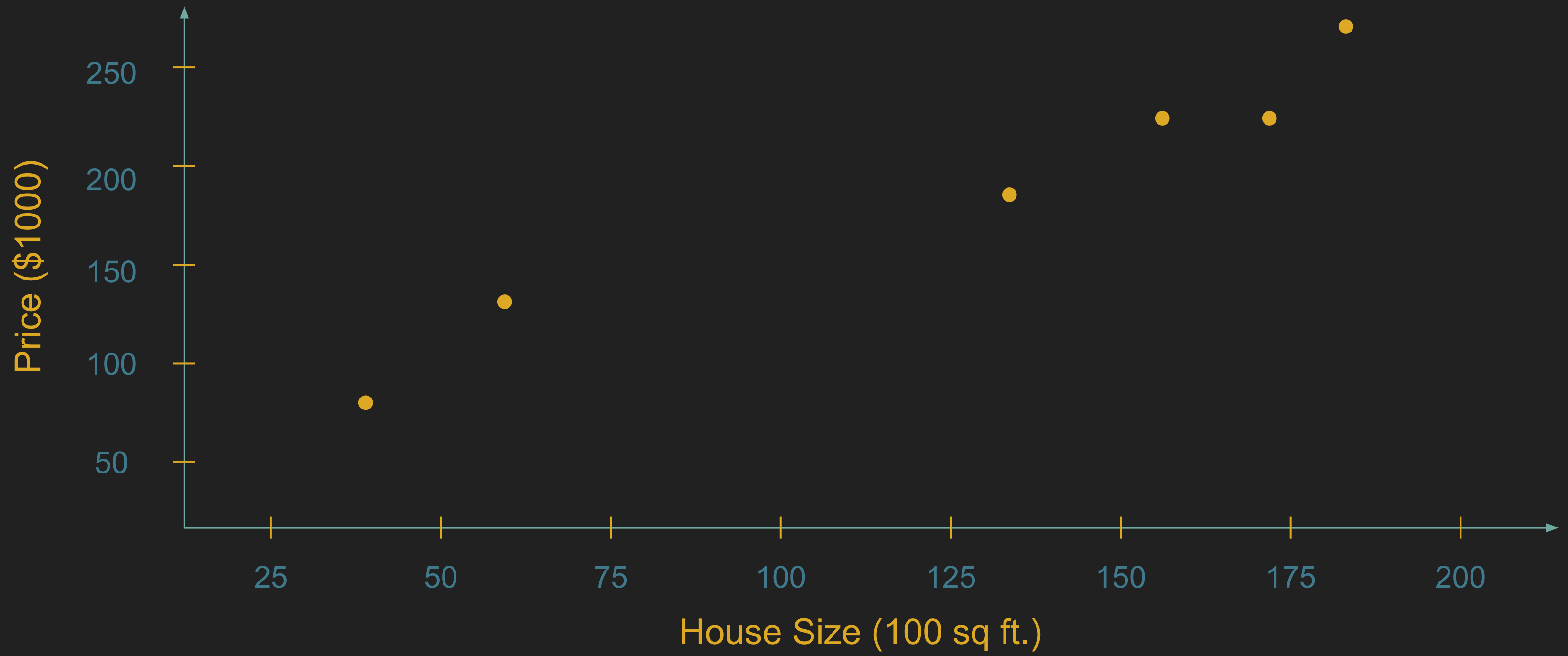


How do we learn from data?

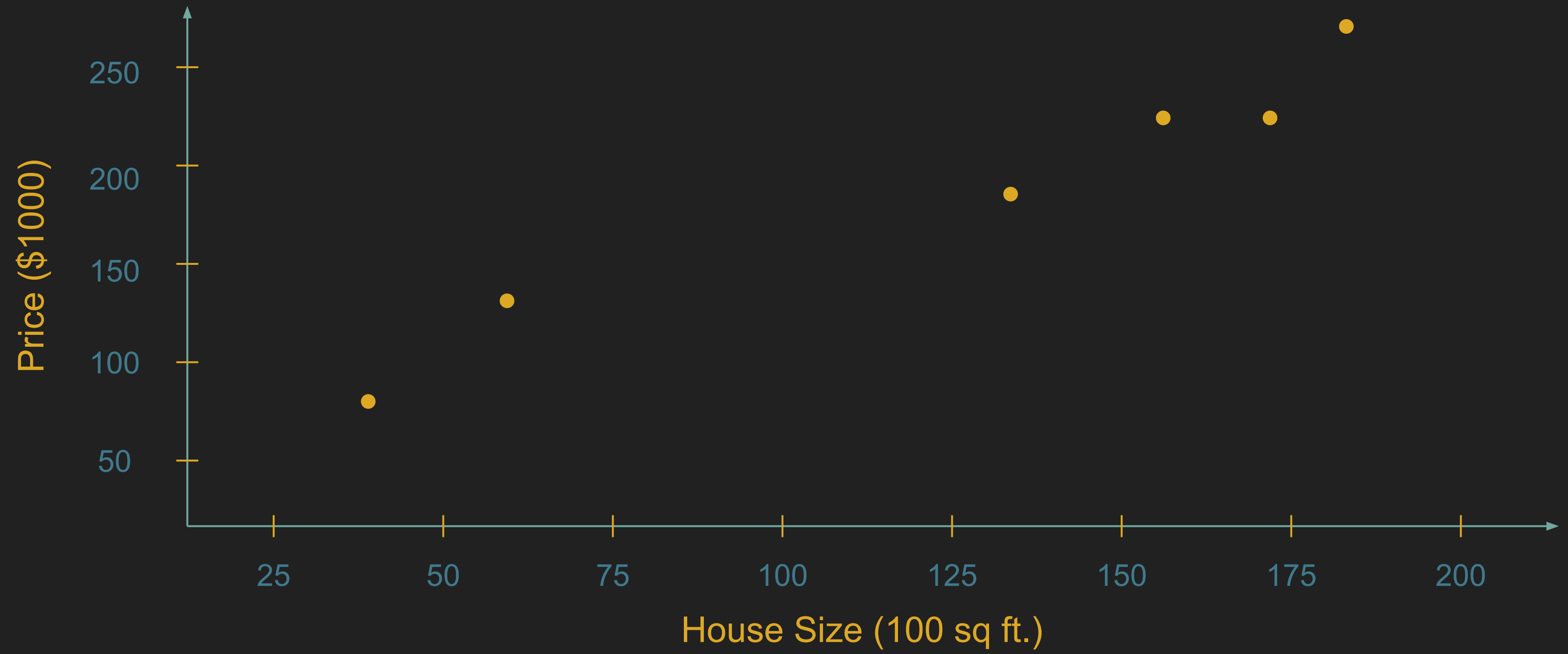
Say we have data points associated with the following graph below.



Is there a trend in this data? If so, what is it?

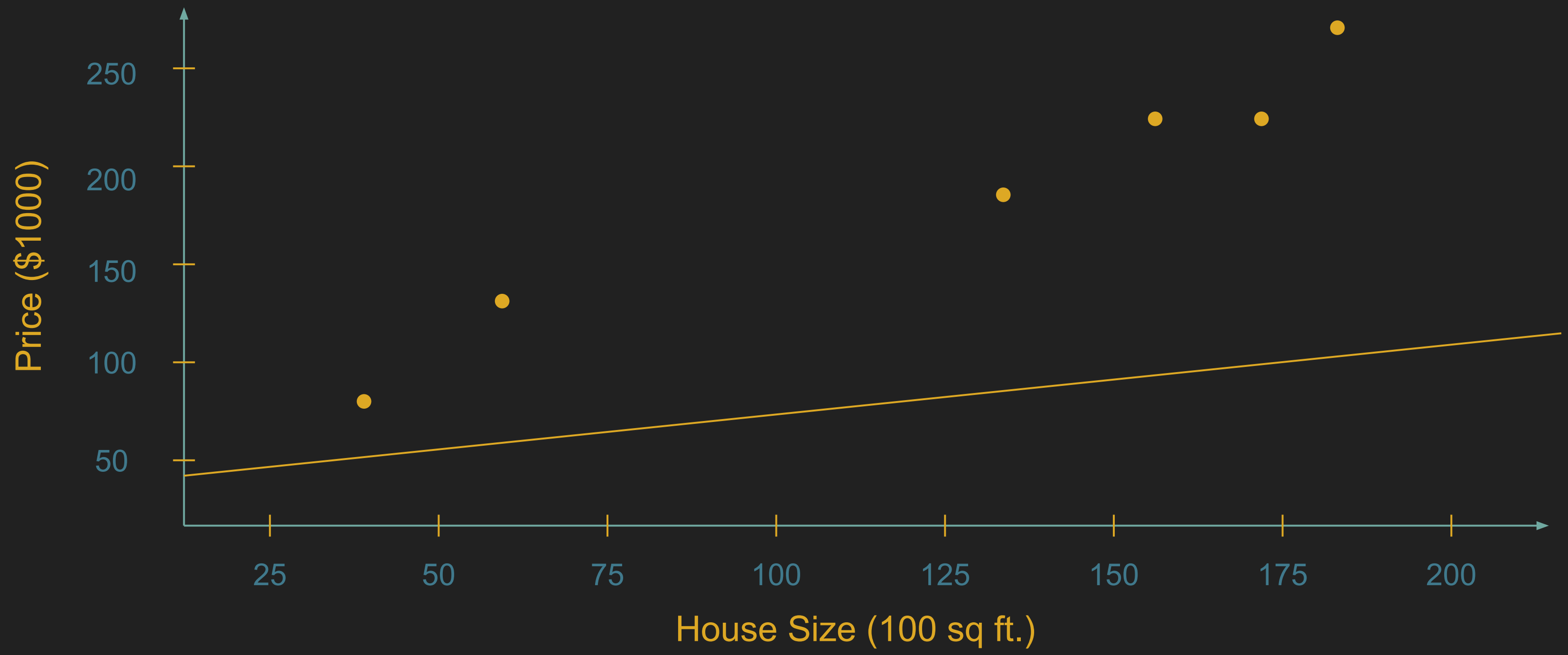


How might a computer learn this trend from these data points?

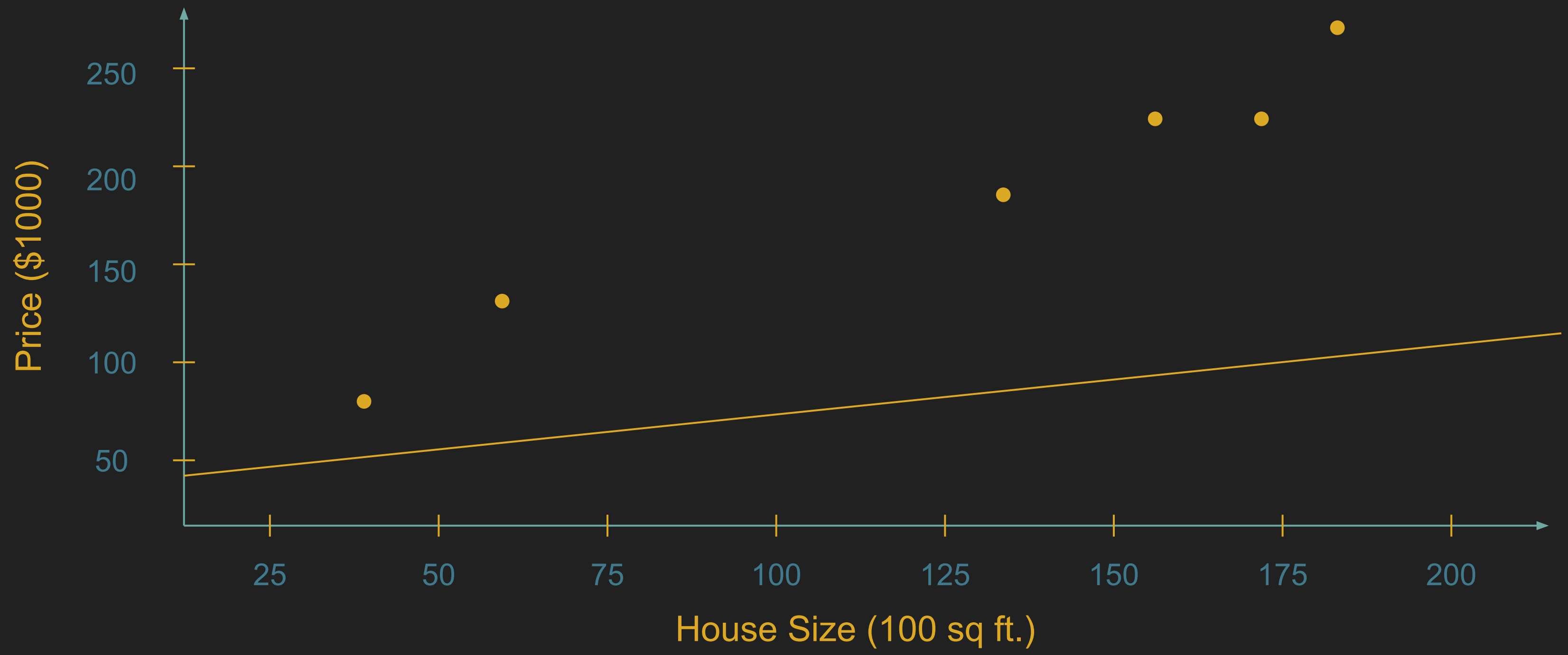


Least Squares Linear Regression (LSLR)

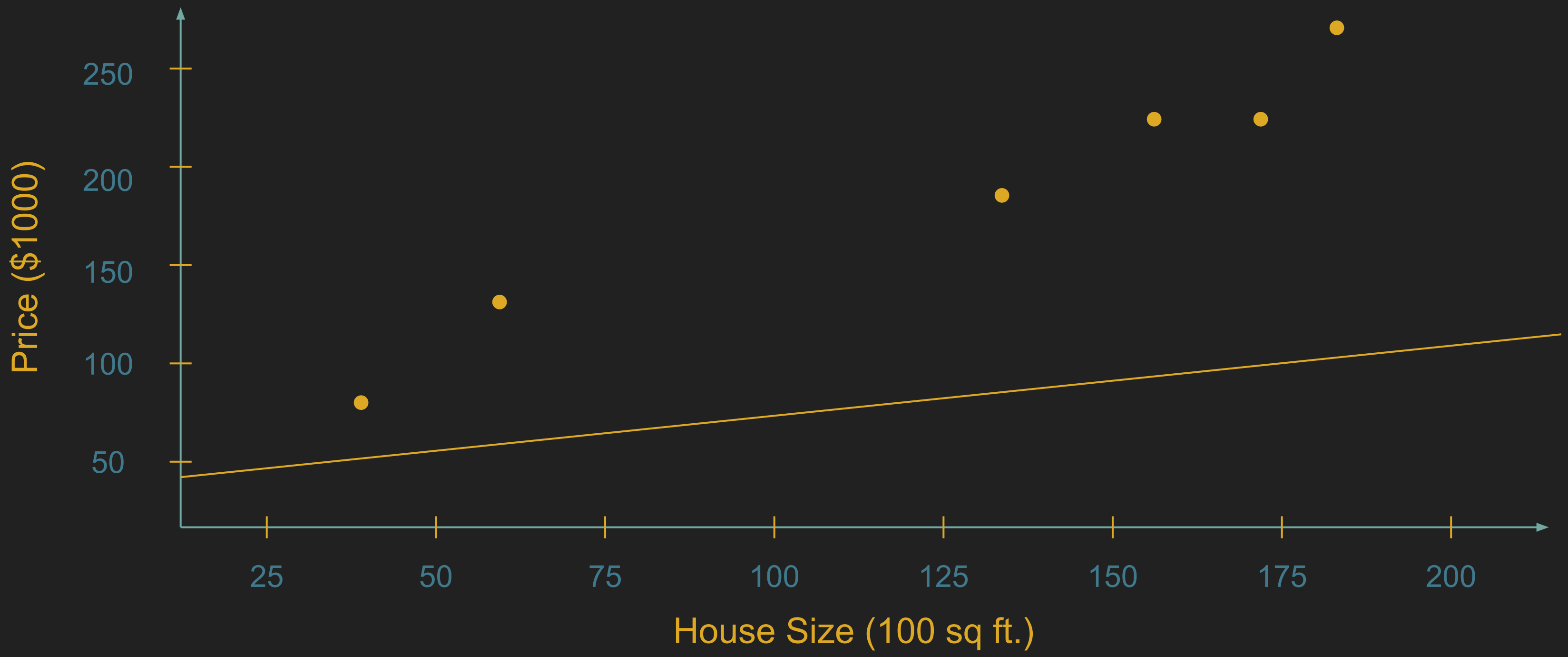
First, the computer makes a guess.



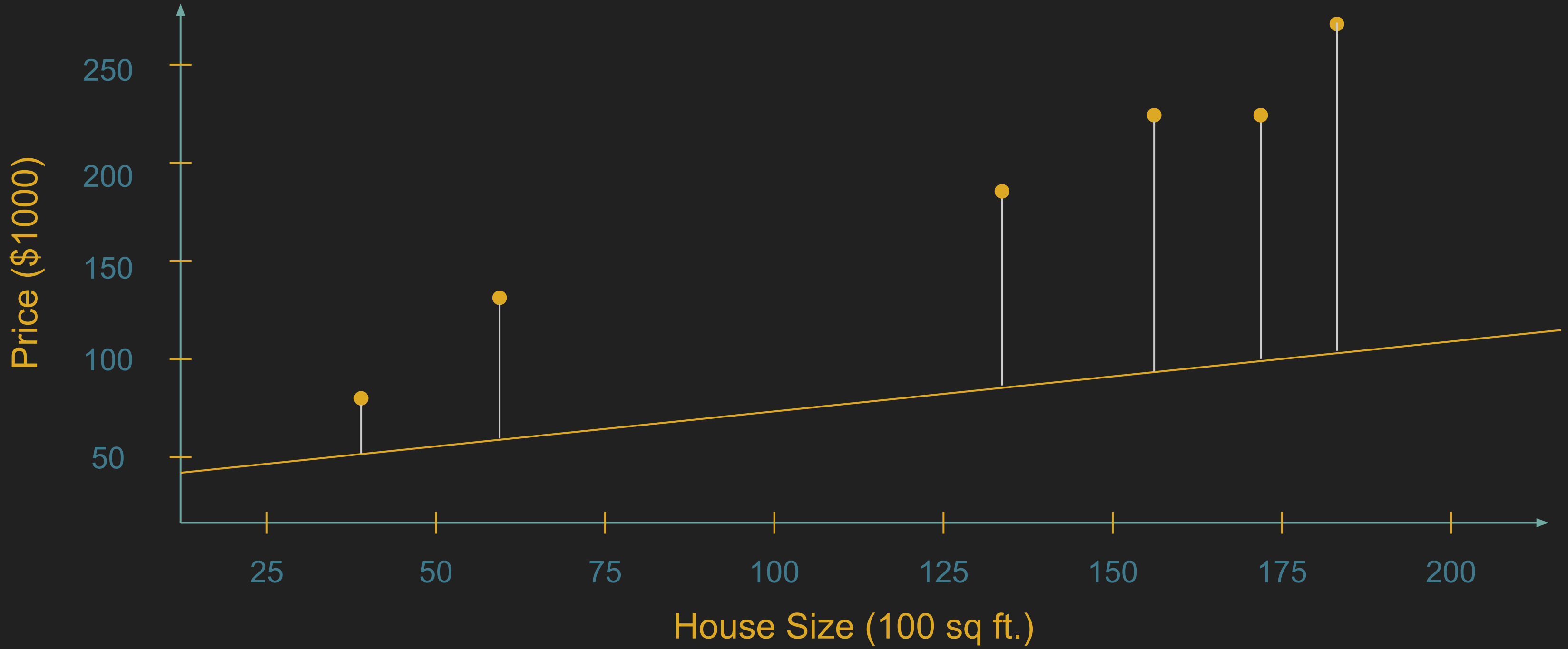
Then it checks its error metric; it wants to know how bad its guess performs with the given data.



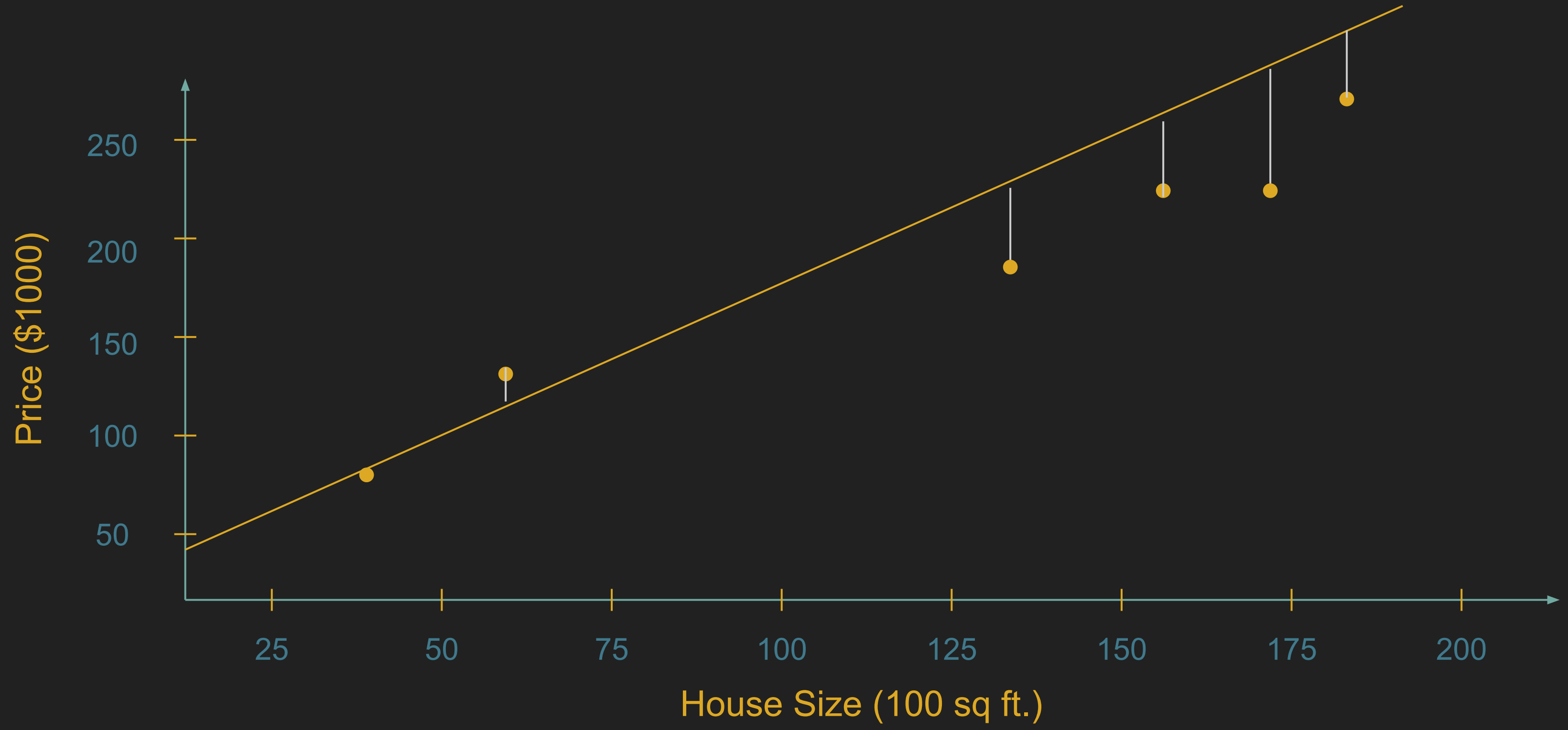
What could the error metric be?



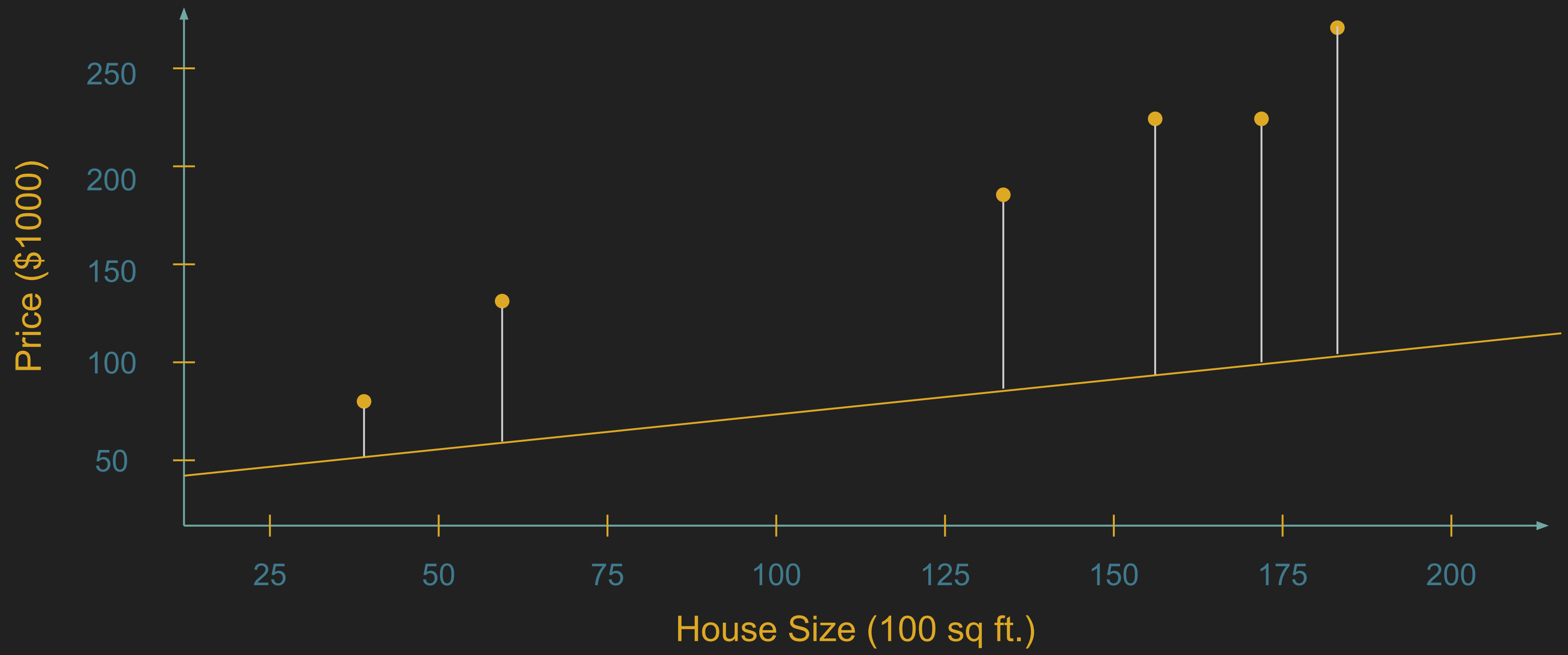
A first guess may be distance from each data point.



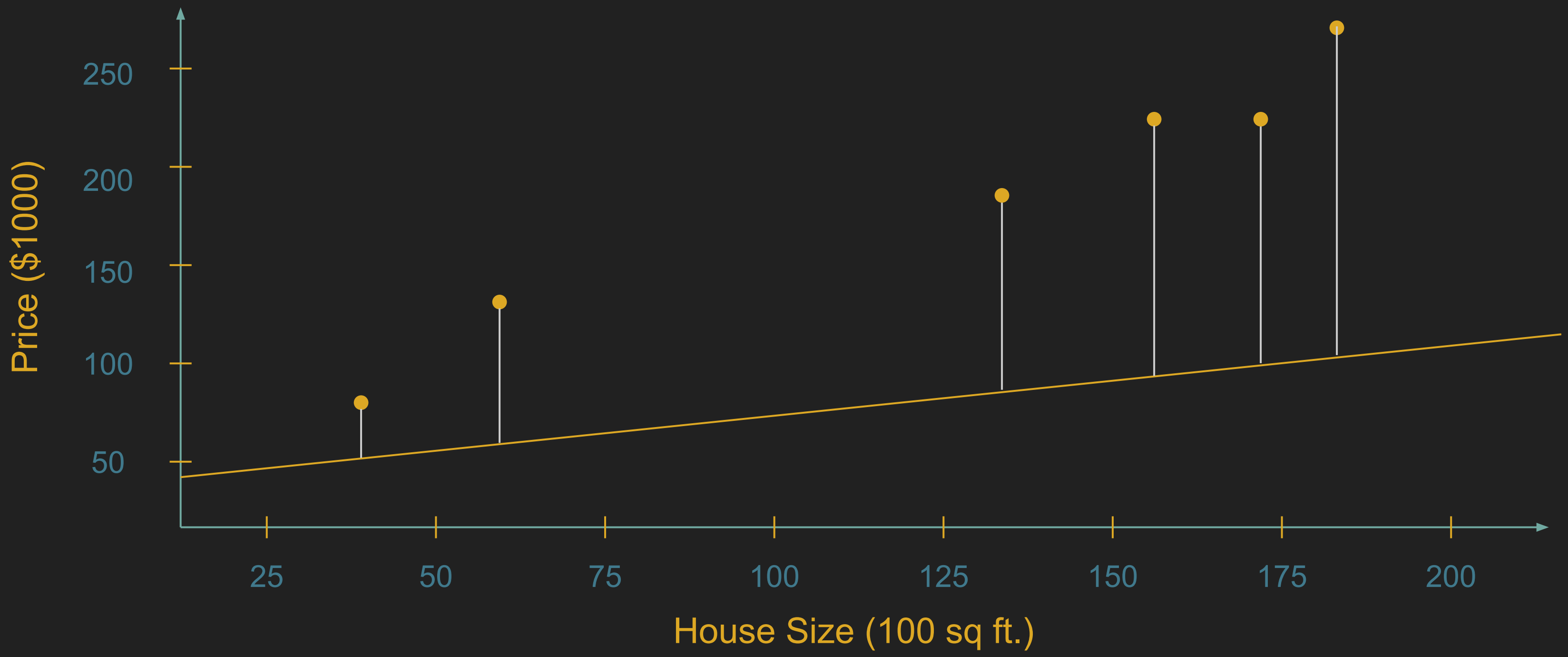
How would we account for negative distances?



We use the mean squared distance of a data point from the guess.

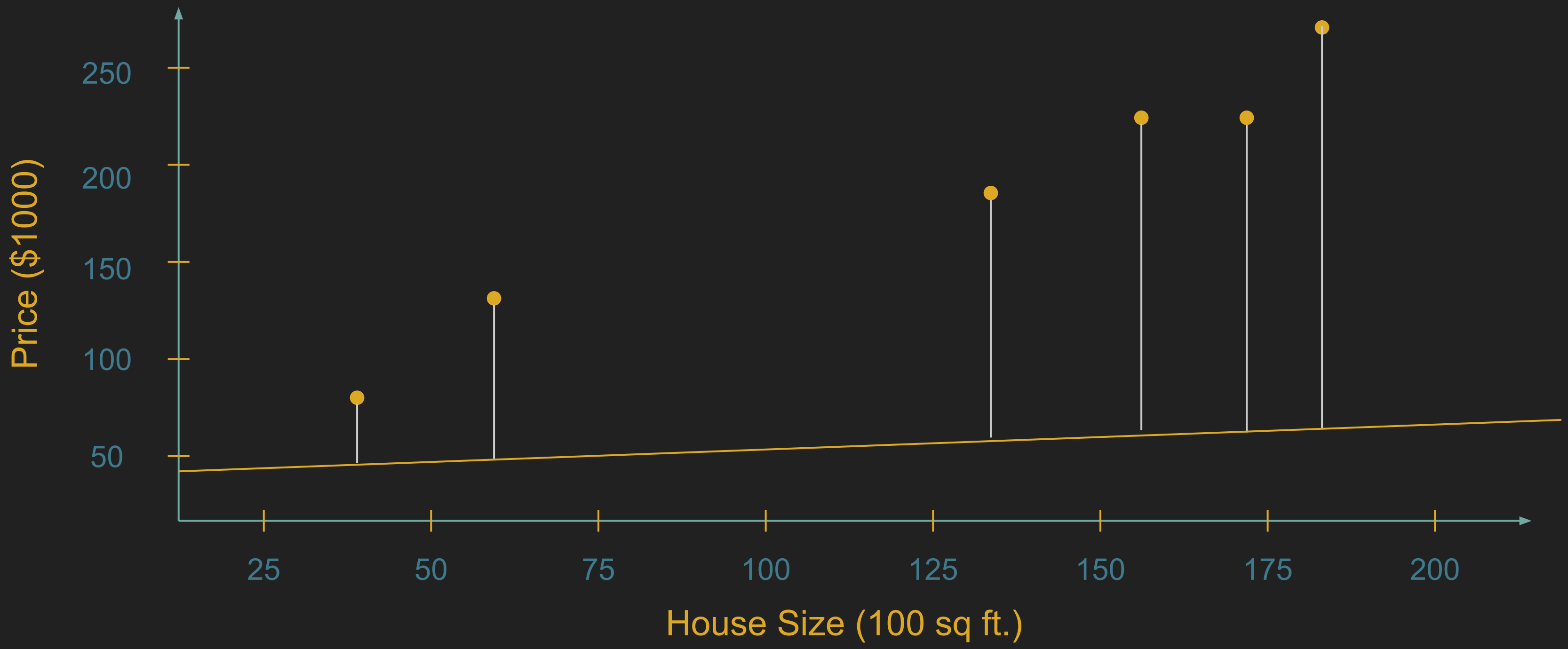


Guess 1


250
200
150
100
50
Price ($1000)
25
50
75
100
125
150
175
200
House Size (100 sq ft.)

Guess 1 ______________________

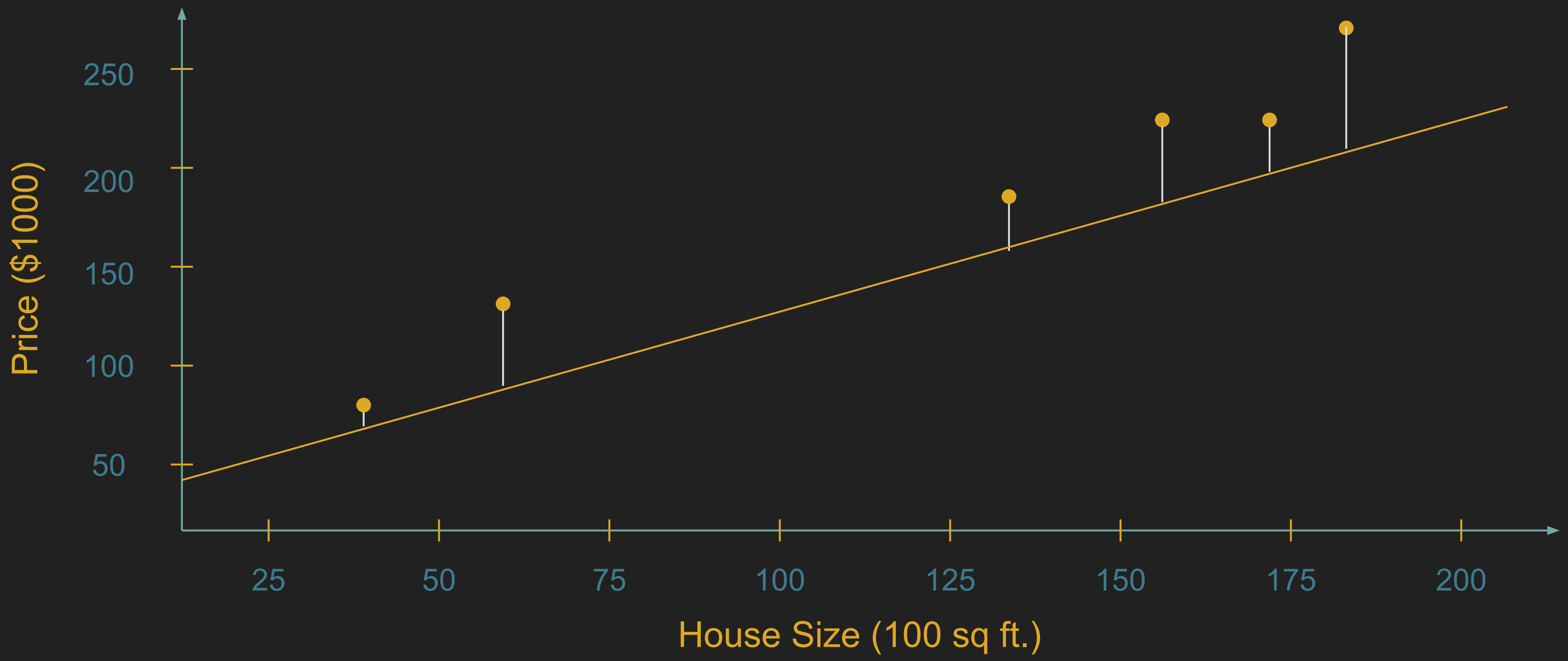
Guess 2 ______________________



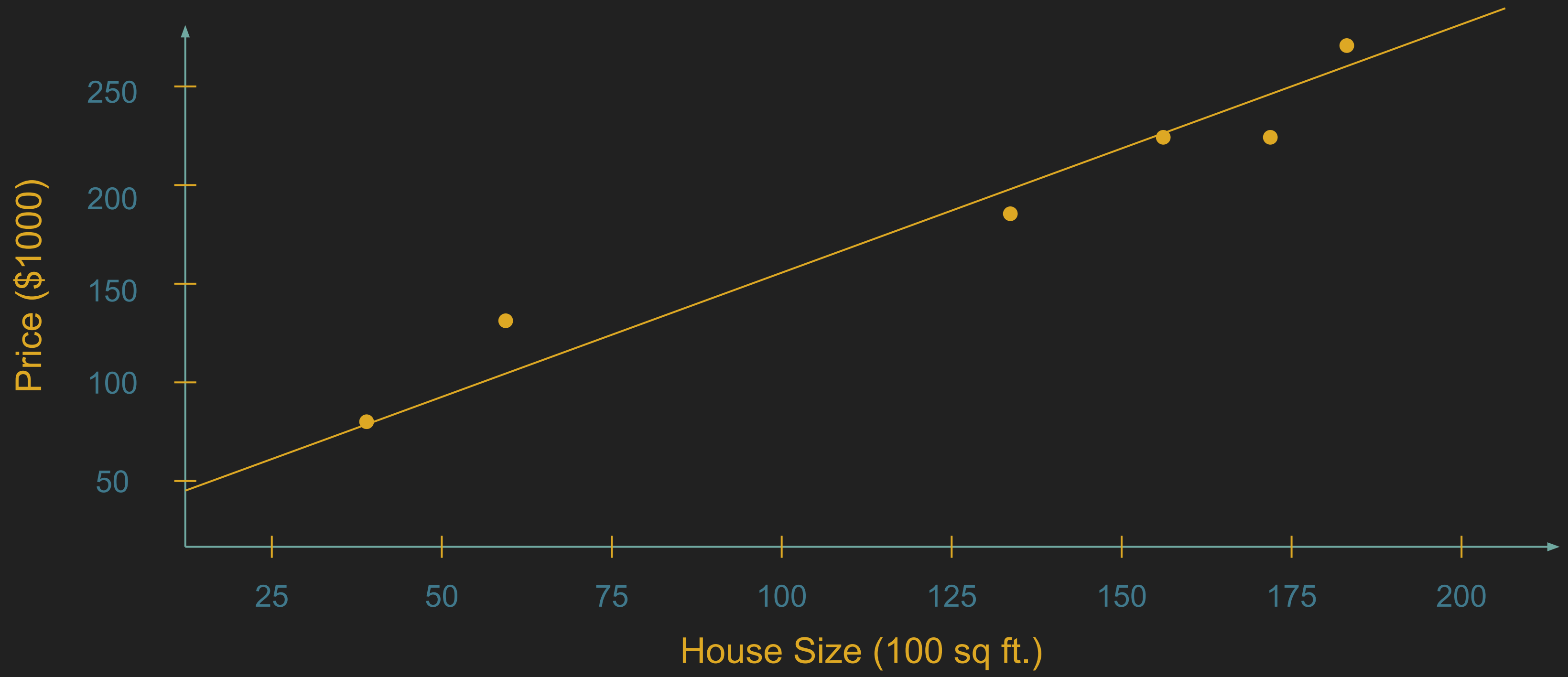
Guess 1

Guess 2

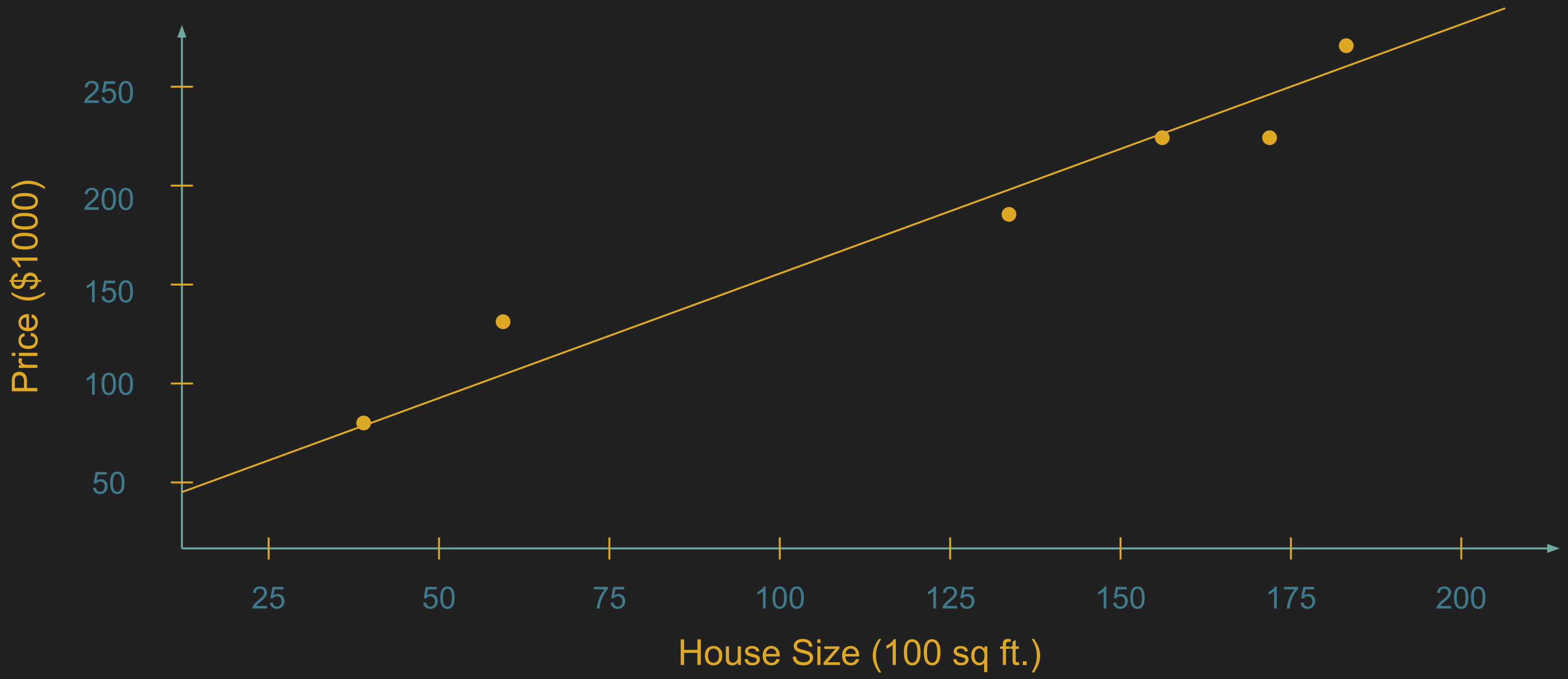
Guess 3



The algorithm will make guesses, calculate the error for each data point at each guess, and try to minimize it until it reaches the minimum mean square error.



This is known as the training phase.



The algorithm is programmed to decrease the mean square error over many iterations.

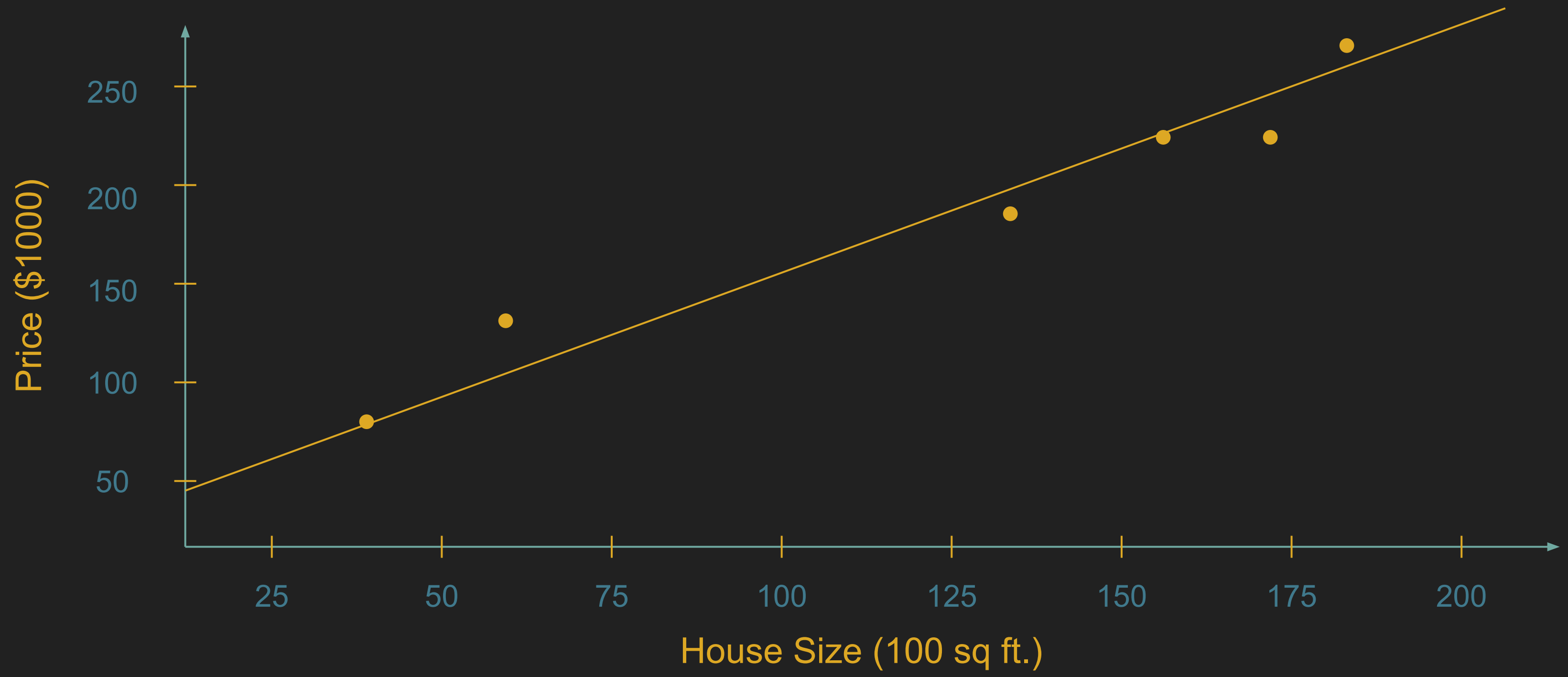
This is a simple example of gradient descent.

What is learned in LSLR?

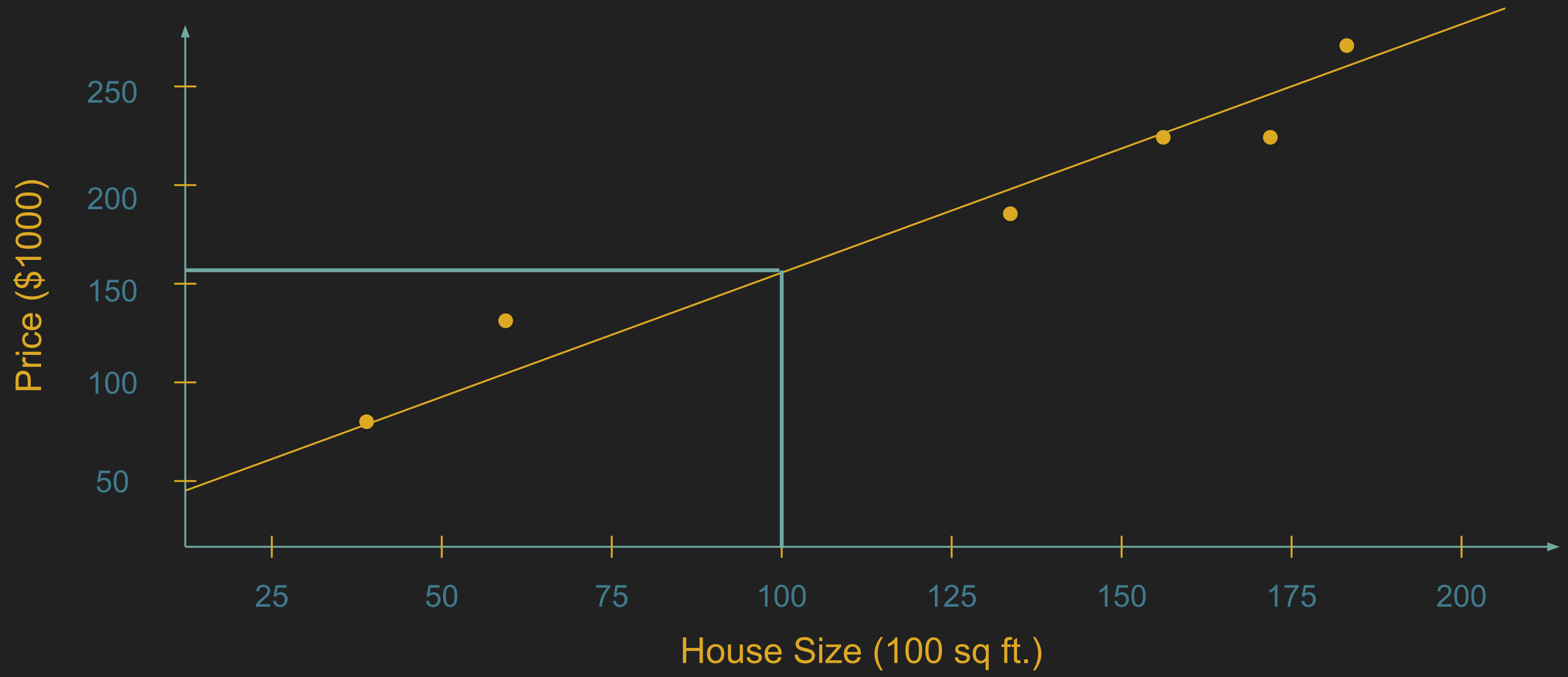
How does it do this?

What type of learning is this?

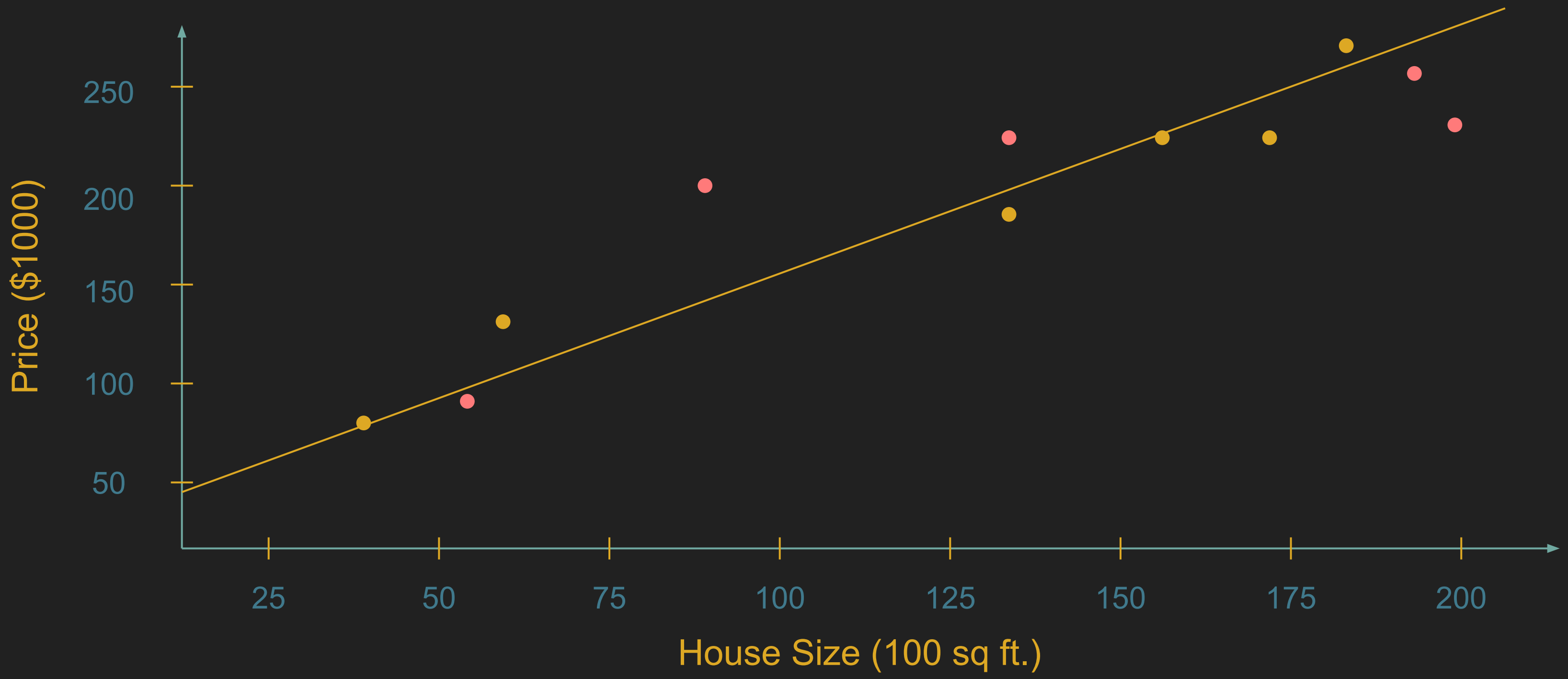
Now, we can use the learned parameters to estimate the price of a house size where we don't have an observation for that house size.



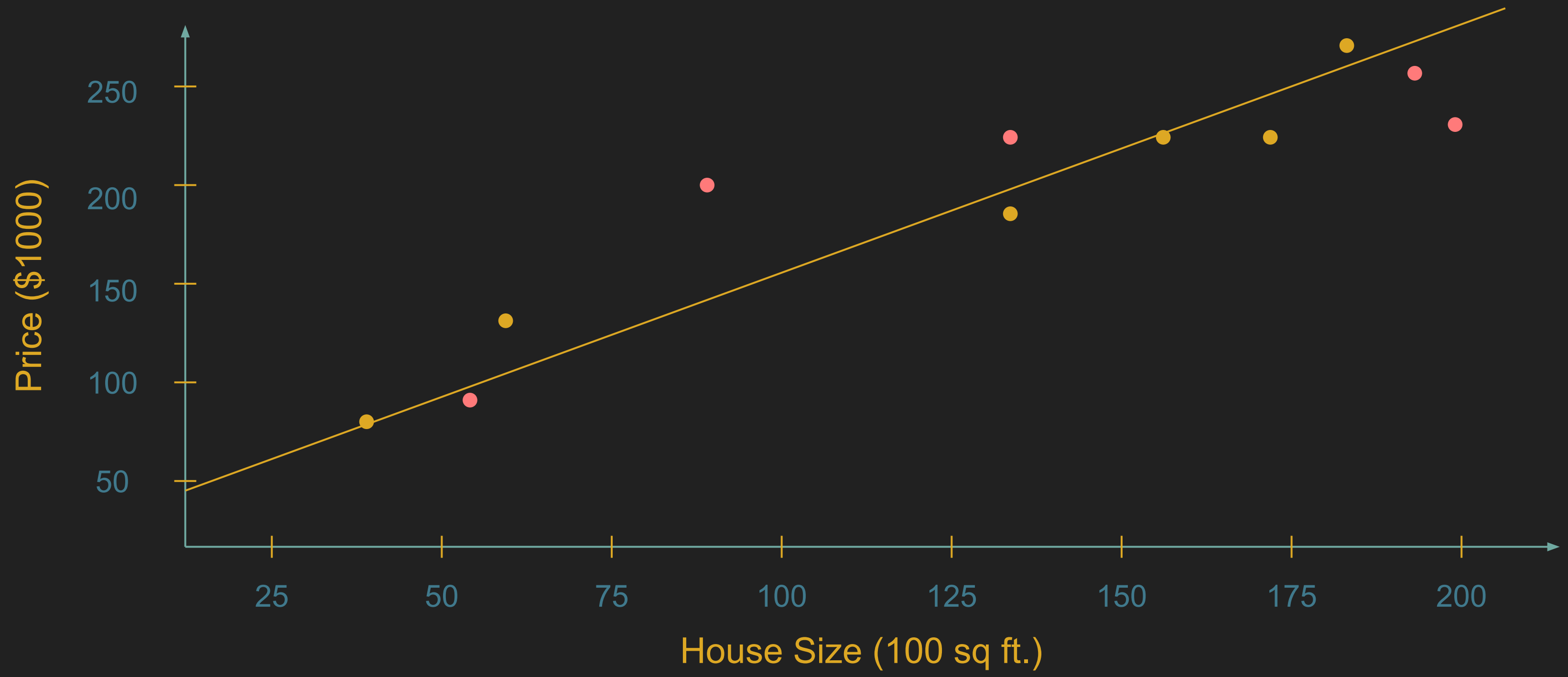
Now, we can use the learned parameters to estimate the price of a house size where we don’t have an observation for that house size.



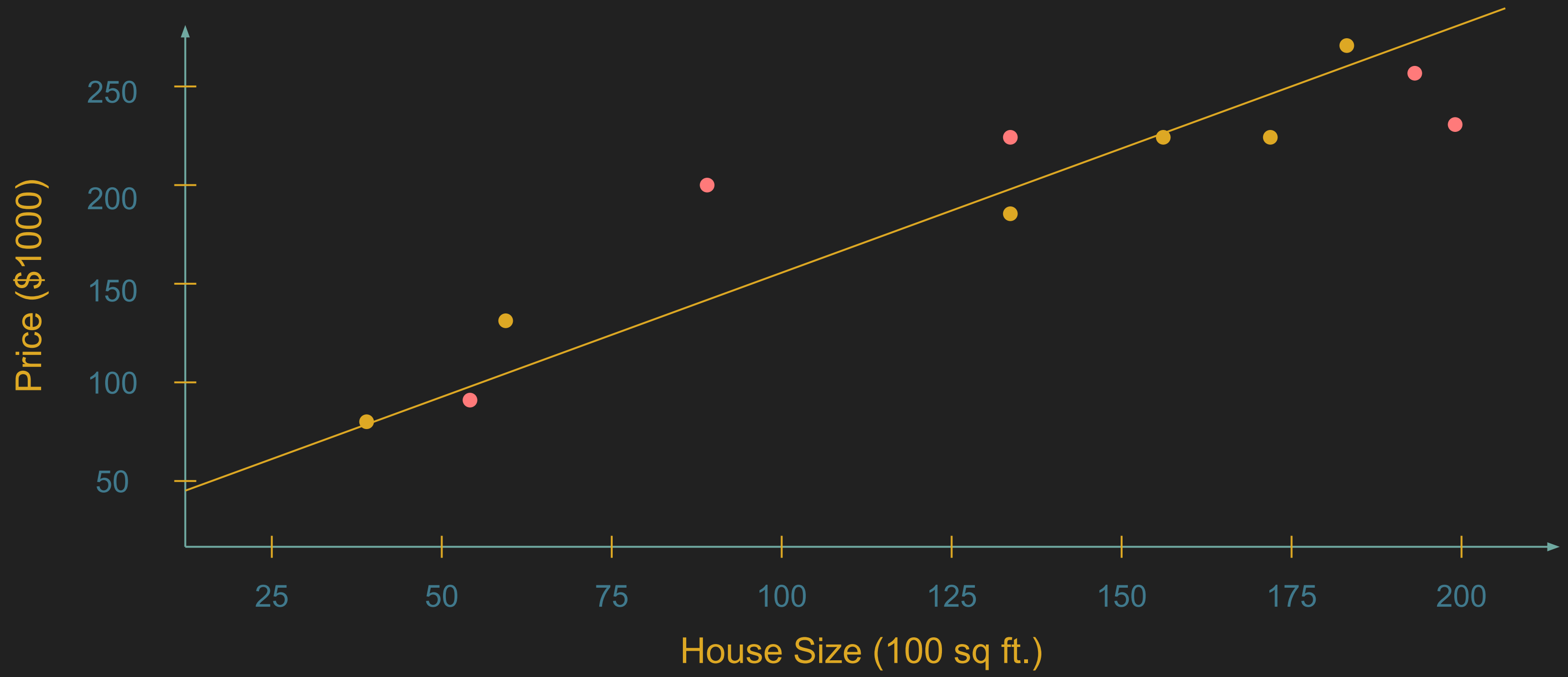
What if we would like to determine whether or not our model was a good model?



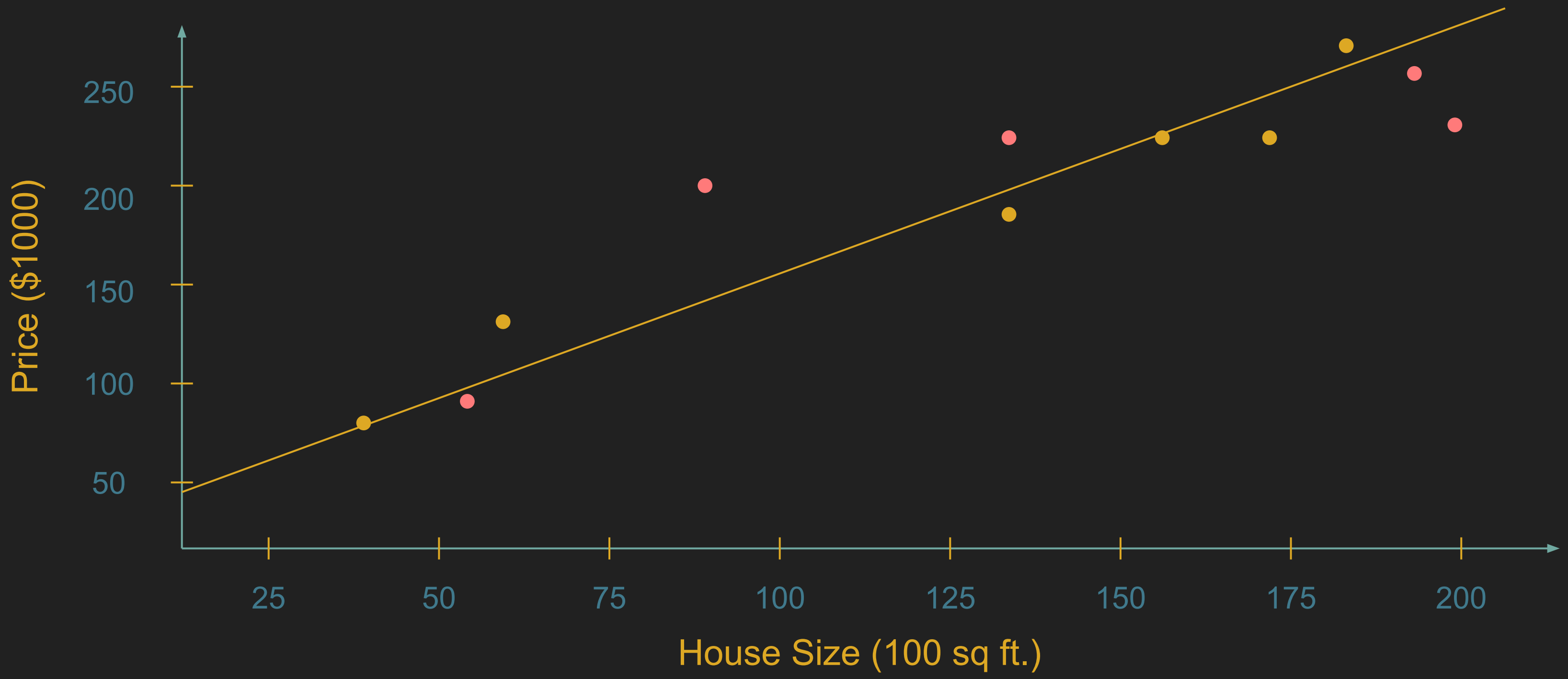
We can check to see how good our parameter fit is by checking other data observations that we may have left out of the set.



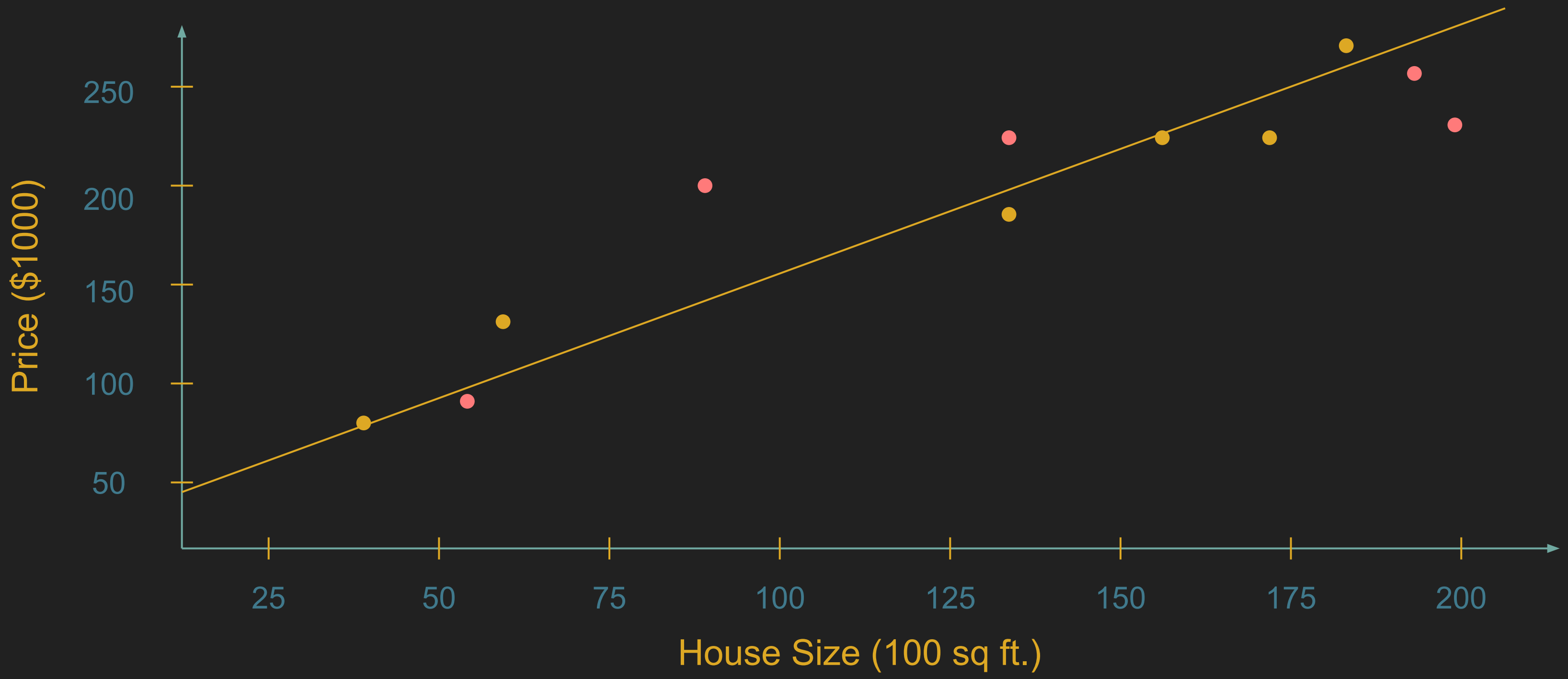
This is called the test phase.



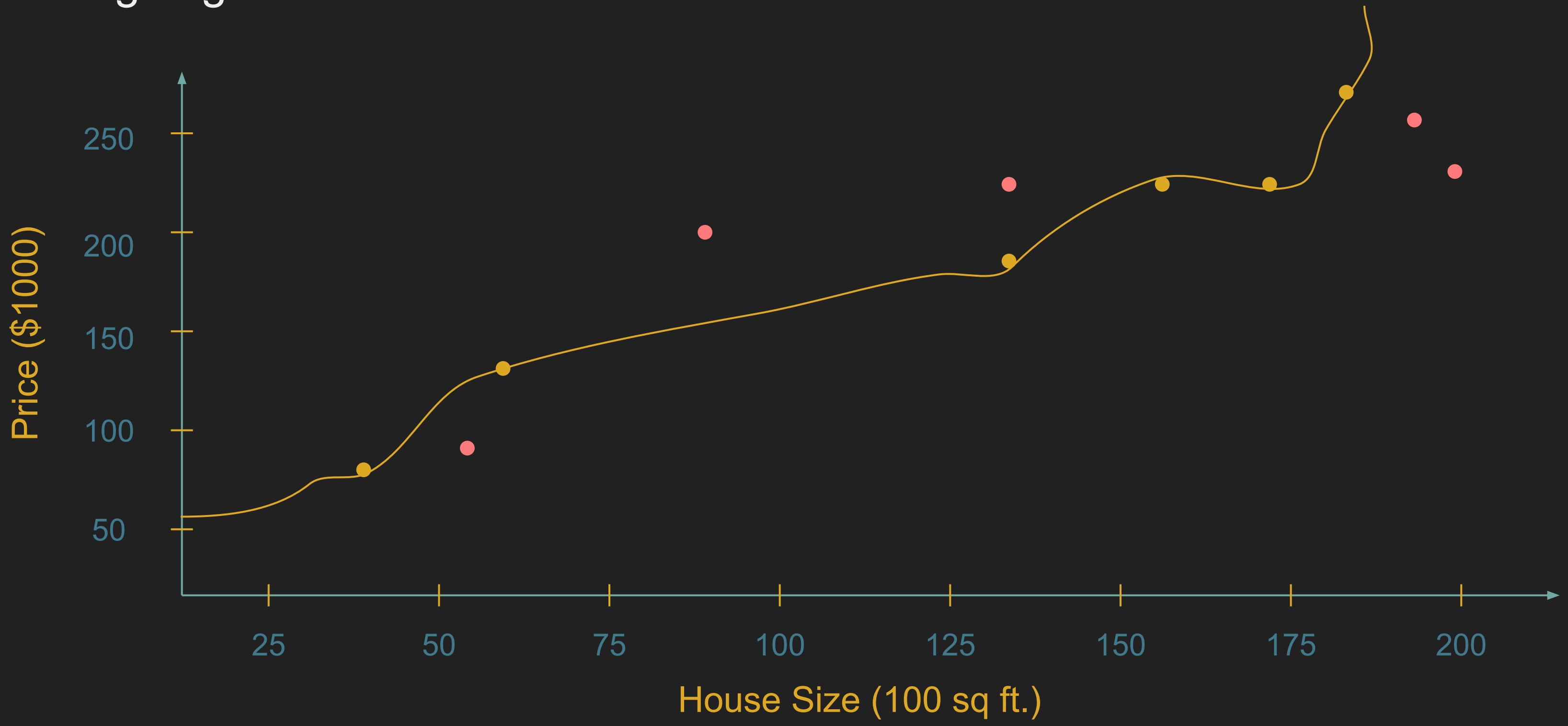
Generalizability can be defined as the extension of research findings and conclusions from a study conducted on a sample set to a larger data set.



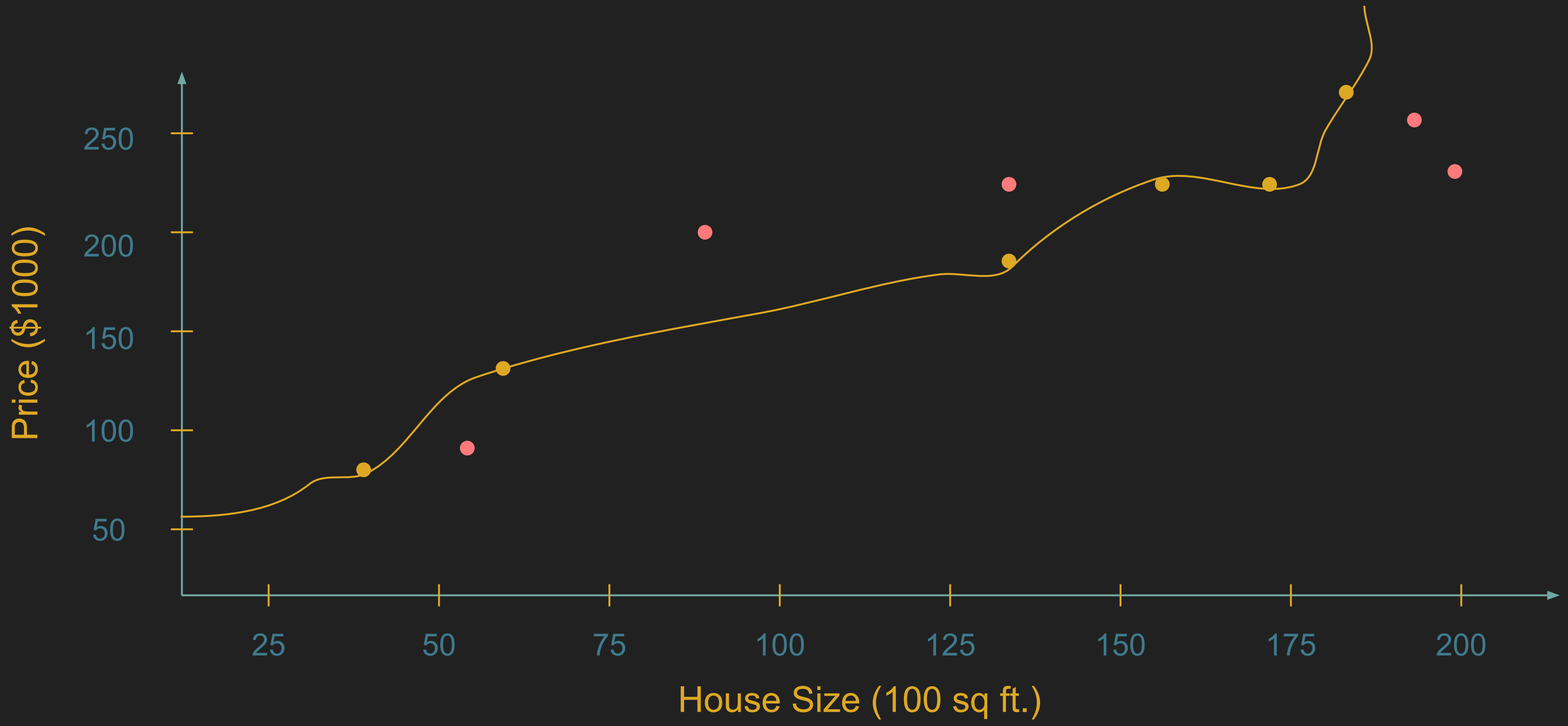
Overfitting is when you train a model and it fits the given data extremely well but fails when new data outside of the training set is introduced.



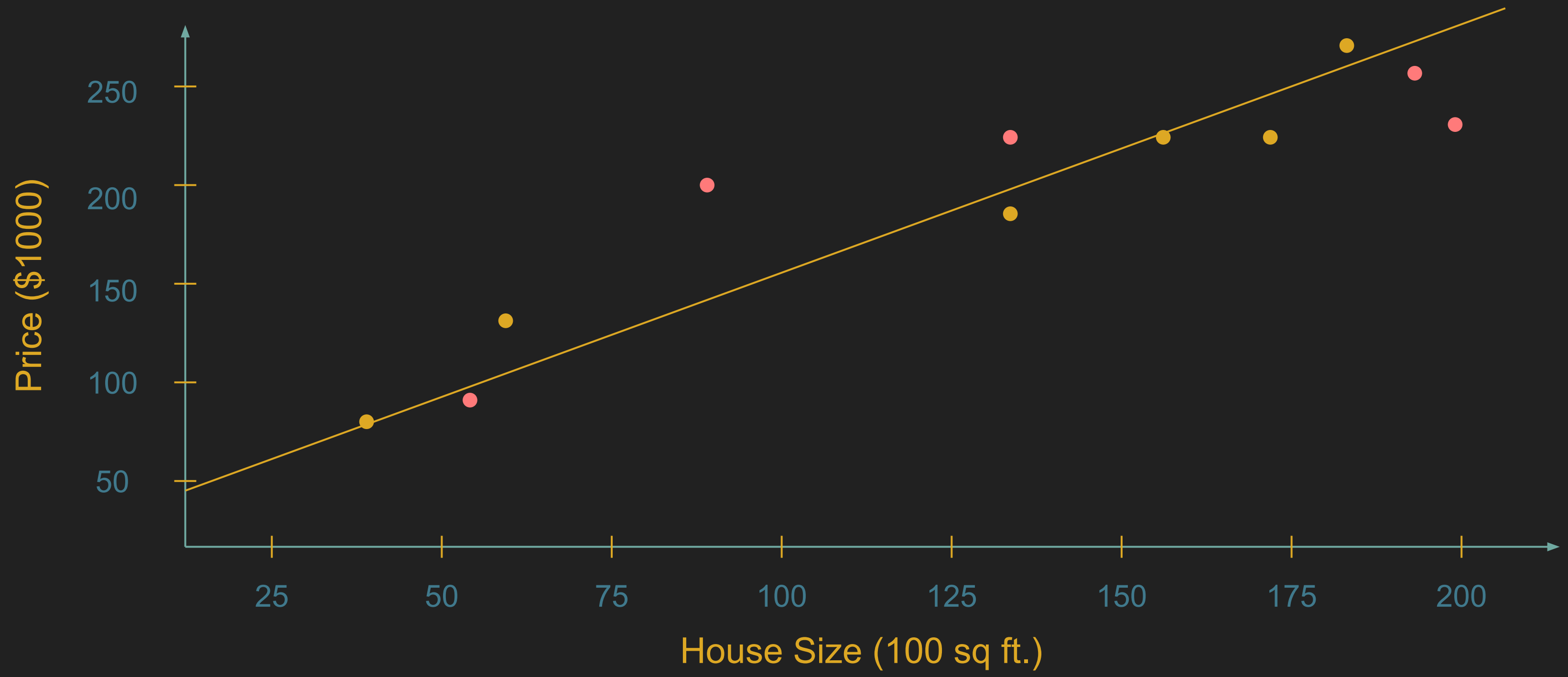
If we relax the constraint that we are training a linear model, we can illustrate what overfitting might look like.



In this case, why do we know that this is less desirable than a linear model?



Generalizability is pivotal for a model to be useful.

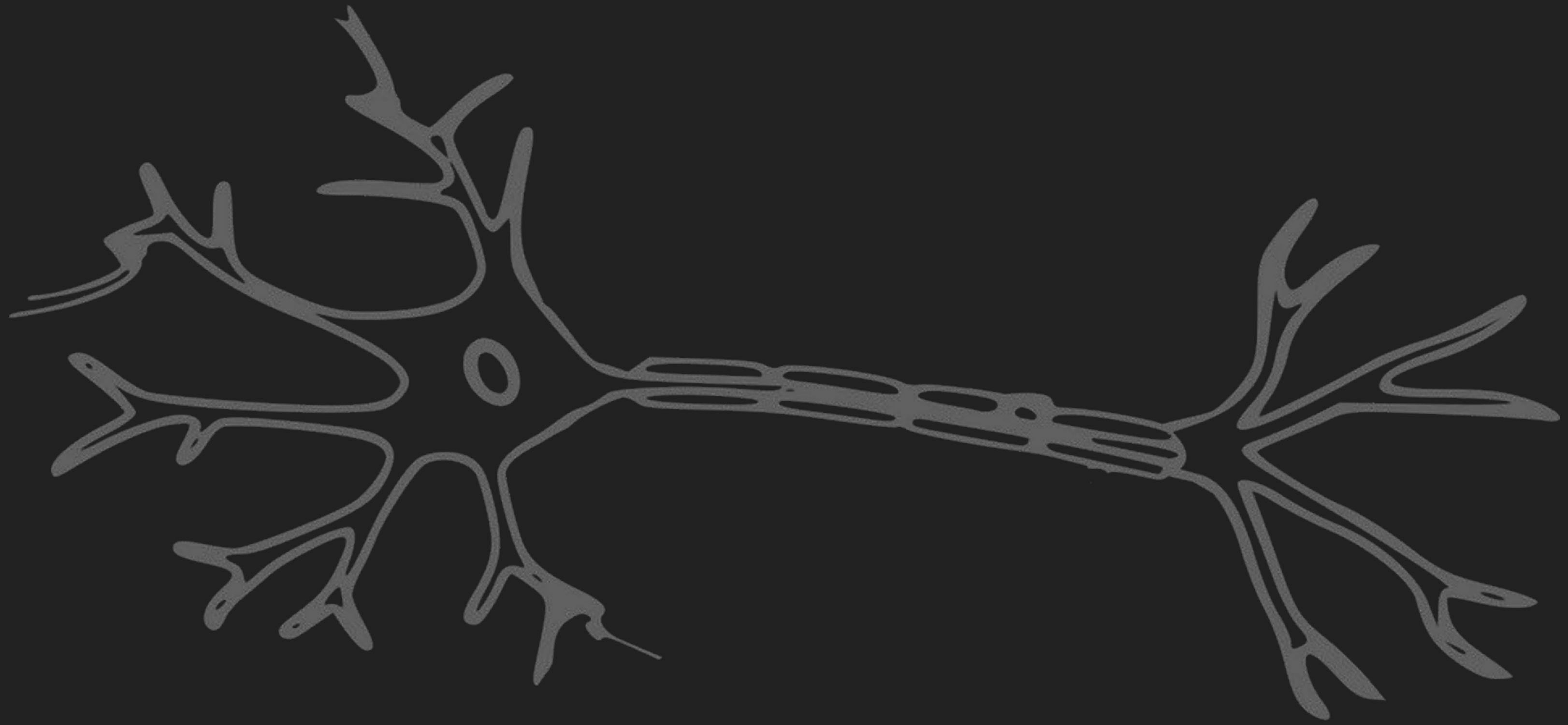


Machine Learning and Biology

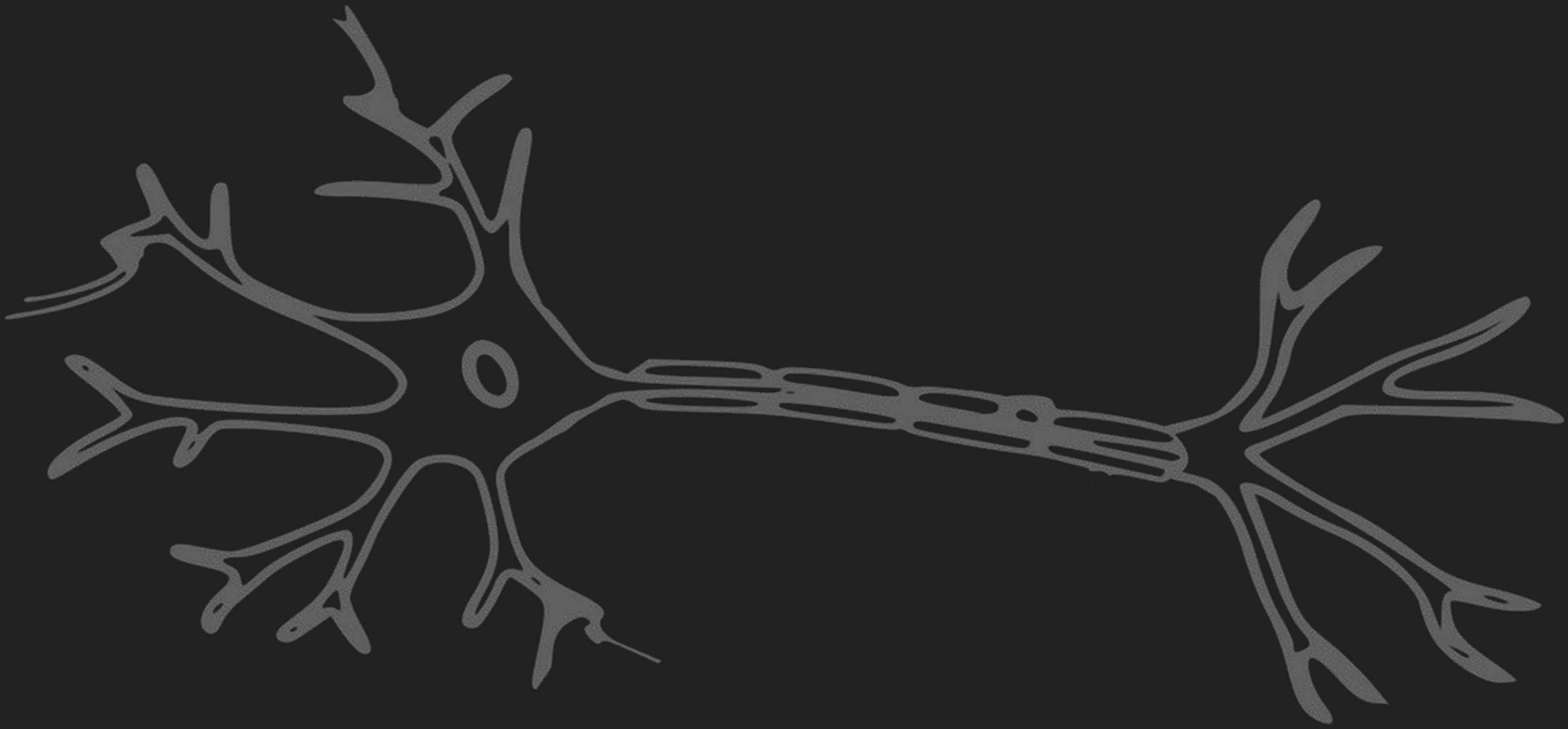
One of my favorite applications of machine learning to neuroscience is that which is done with brain-computer interface research.

How might machine learning be used in this case?

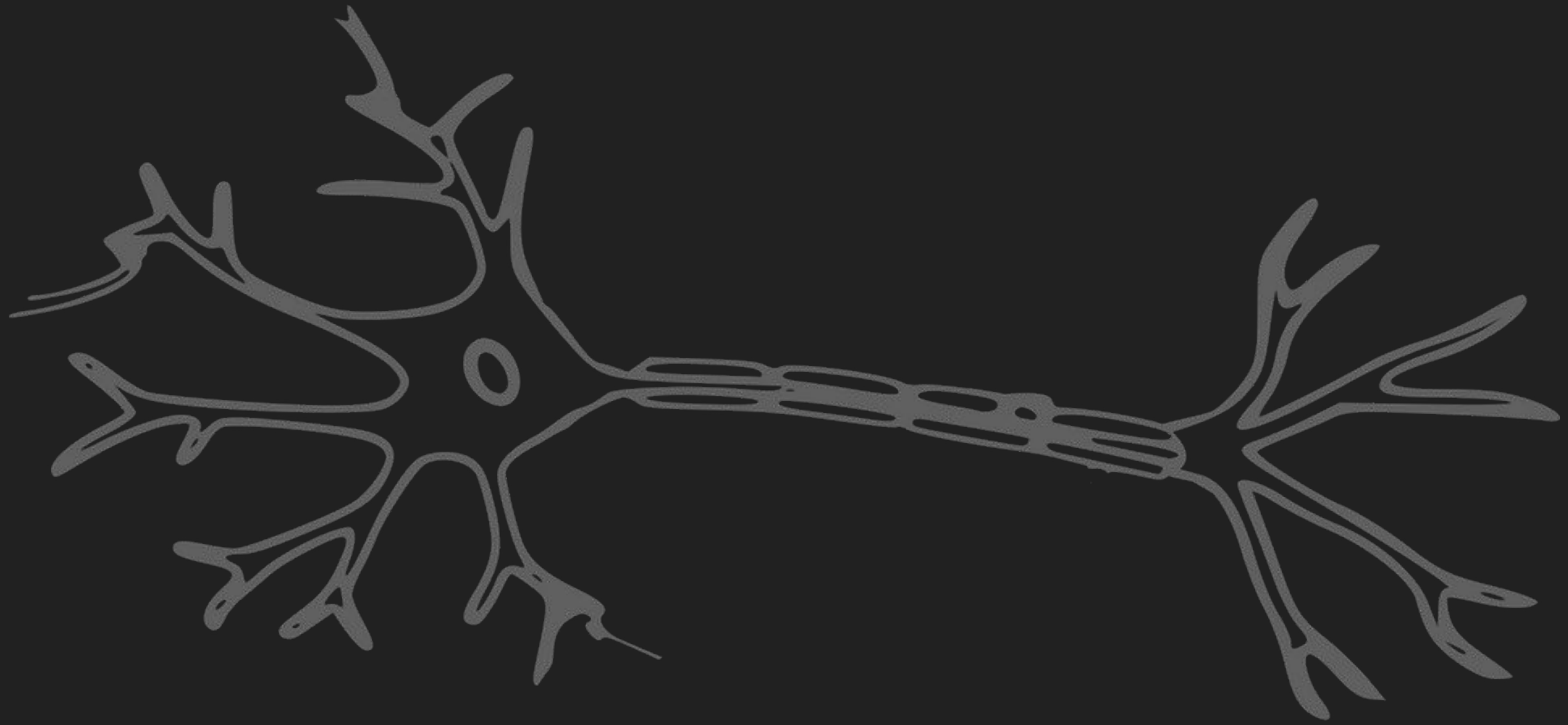
You should have gotten a brief introduction to neuroscience through the previous lectures; what have you learned in these lectures?



So, we have talked a lot about individual neurons and their electrical signals.



In single cell electrical signals, we look at action potentials, so our signals look like spikes.



In this experiment, we are using an MEG, which is a huge machine that measures the magnetic activity of neurons.

Let’s do a bit of review to begin to understanding what it means to read magnetic signals from the brain.

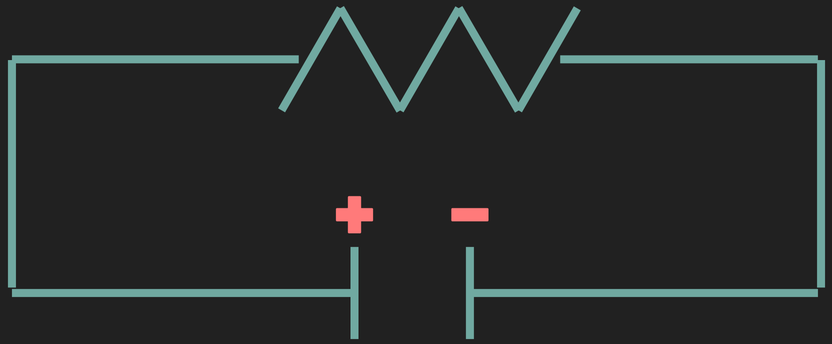
First, let's think about voltage and current.

$$V = IR$$

Voltage = Current * Resistance

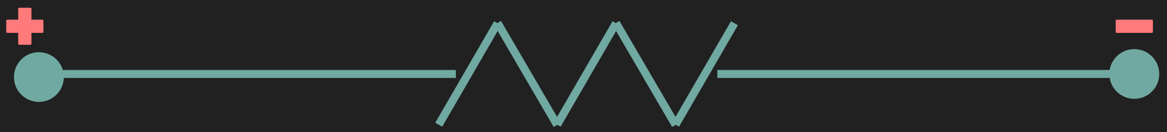
First, let's think about voltage and current.

$$V = IR$$

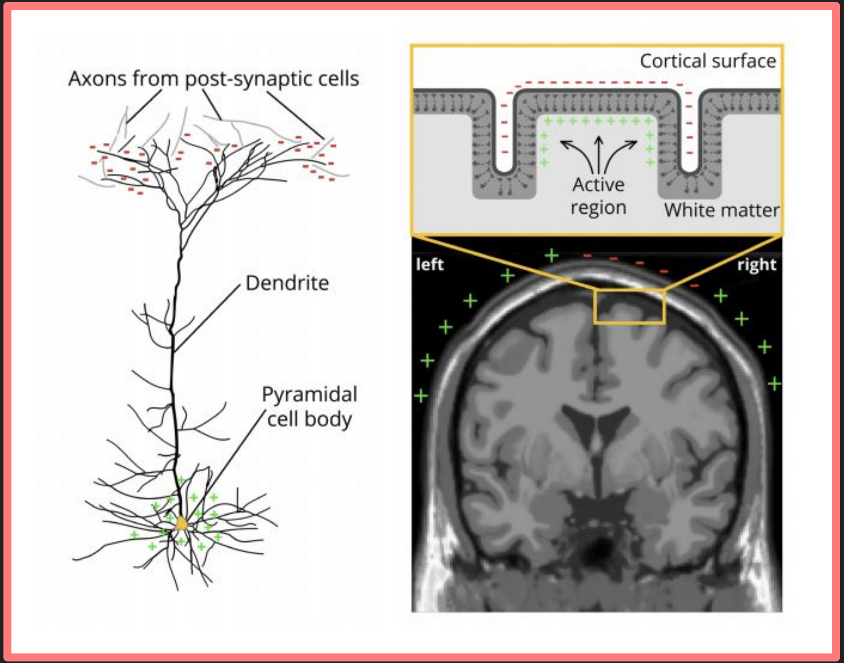


First, let's think about voltage and current.

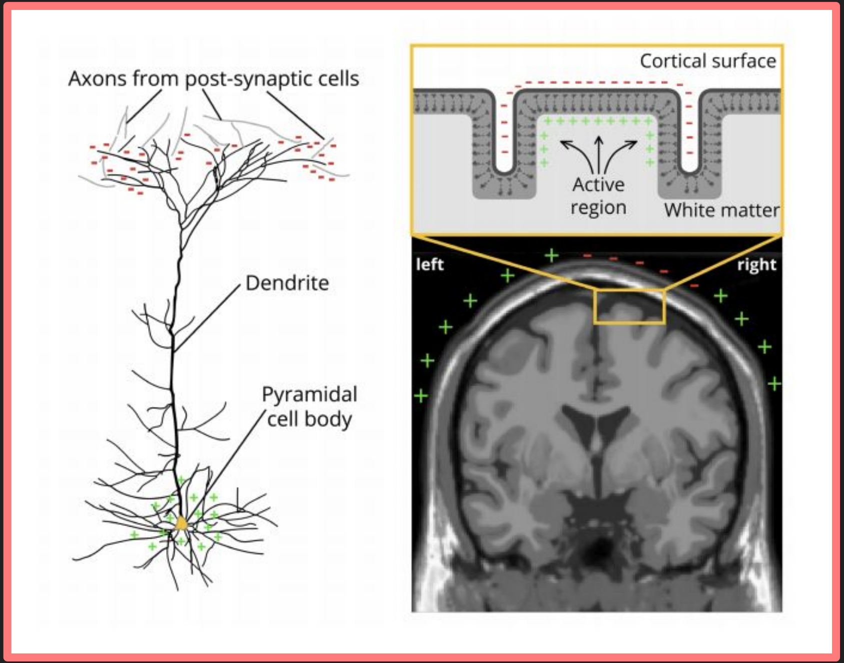
$$V = IR$$



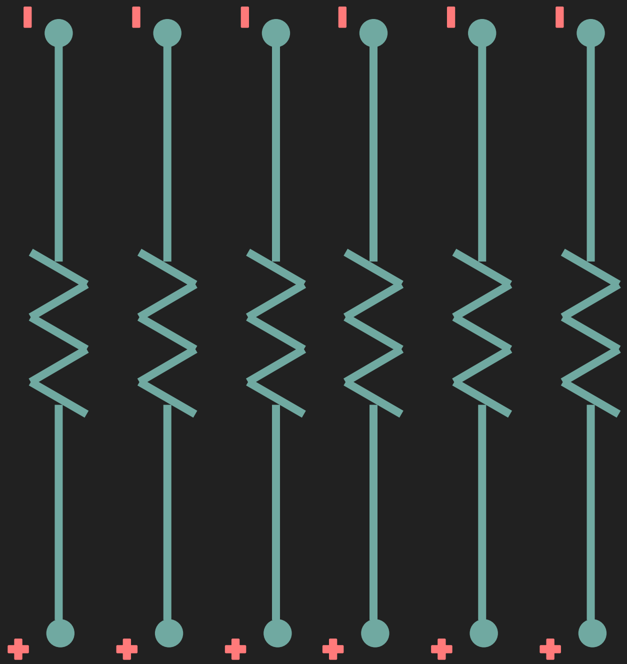
There are voltage sources in the brain which we can look at with noninvasive techniques.



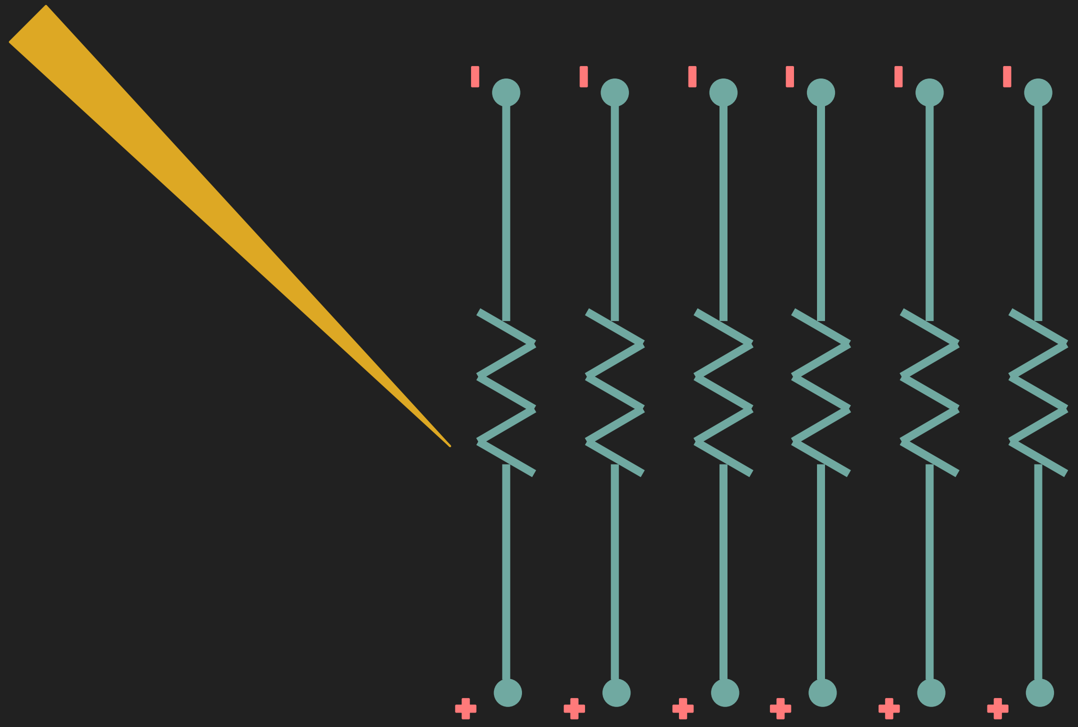
Research indicates that it is primarily the synchronized activity of pyramidal neurons in cortical brain regions which can be measured from the outside.



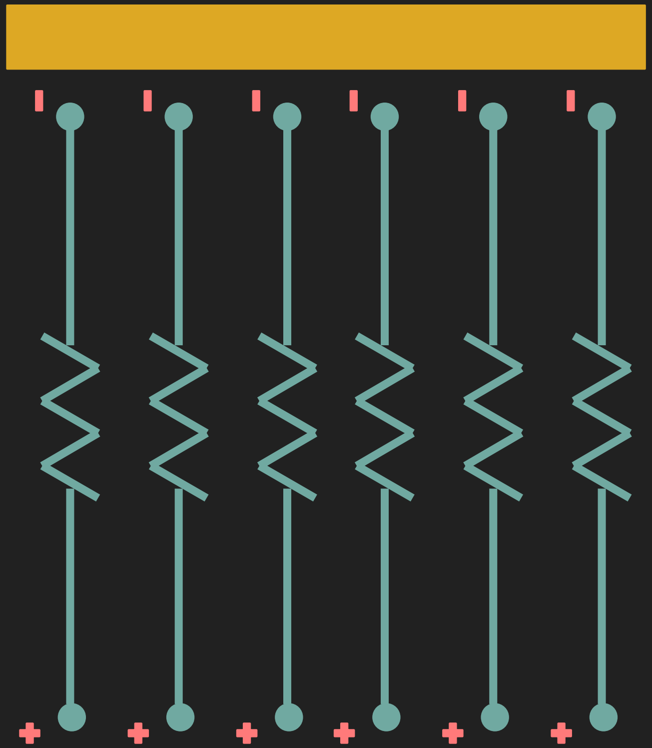
This is because pyramidal neurons are oriented together, and there are a bunch of them we can actually read these synchronous signals through the scalp and bone and things!



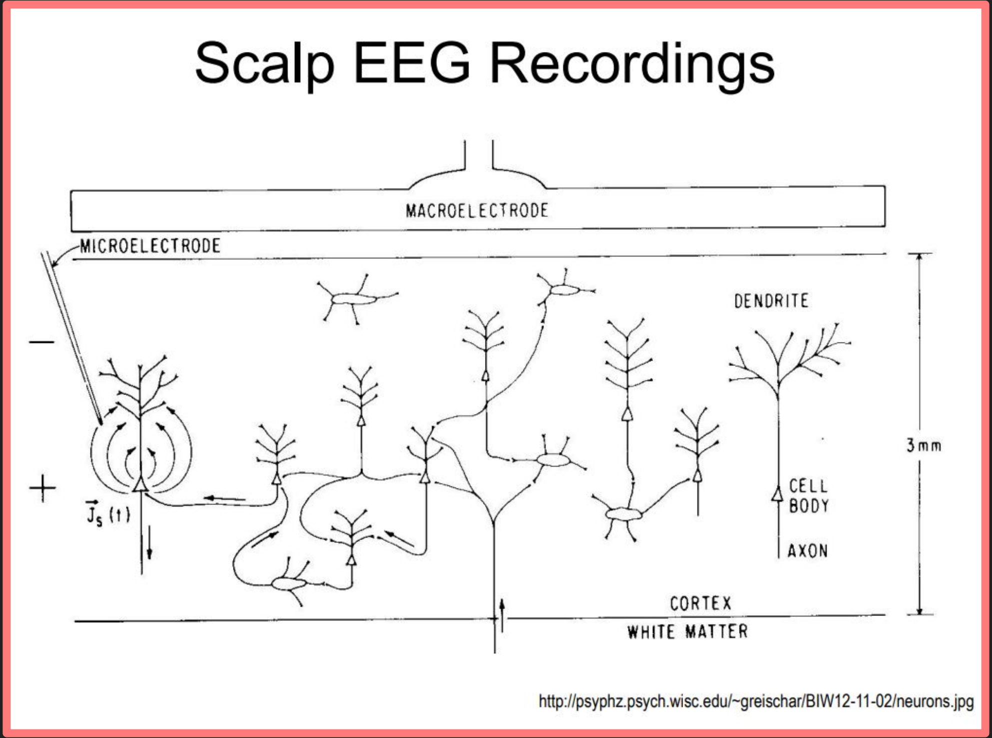
So instead of single neuron recordings, we look at the summed activity of many pyramidal neurons.



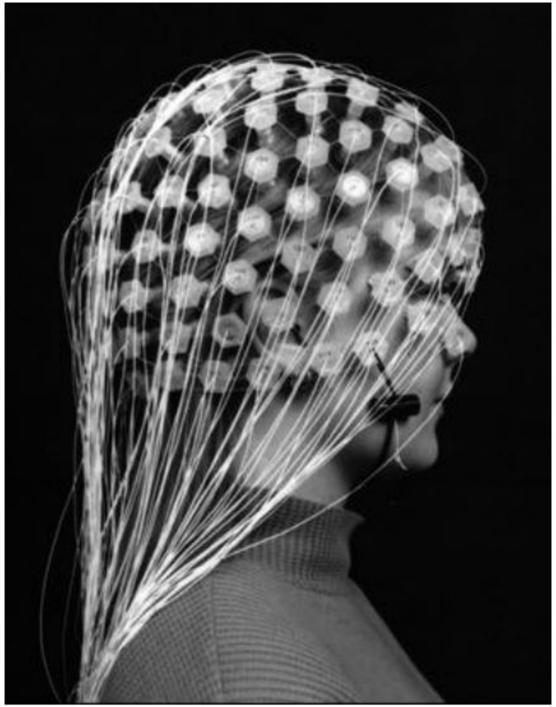
So instead of single neuron recordings, we look at the summed activity of many pyramidal neurons.



So instead of single neuron recordings, we look at the summed activity of many pyramidal neurons.



Electroencephalography (EEG) looks like this!

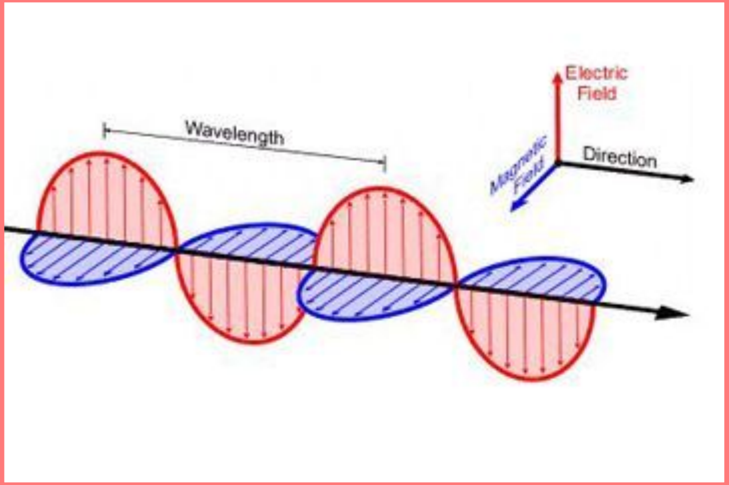


http://www.scholarpedia.org/wiki/images/1/10/Electroencephalogram_figHead.jpg

http://www.thecnnh.org/ACEC/images/jimmysmile_phan.jpg

Where the heck does a magnetic field come in?

Maxwell's equations: changing magnetic fields produce electric fields



$$\nabla \cdot \mathbf{E} = \frac{\rho}{\varepsilon_0}$$

$$\nabla \cdot \mathbf{B} = 0$$

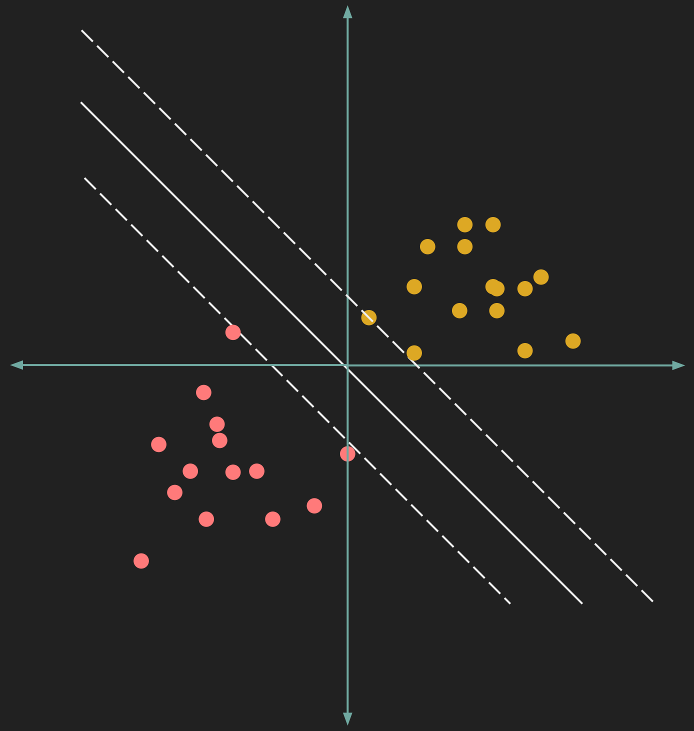
$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}$$

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{j} + \frac{1}{c^2}\frac{\partial \mathbf{E}}{\partial t}$$

SVM

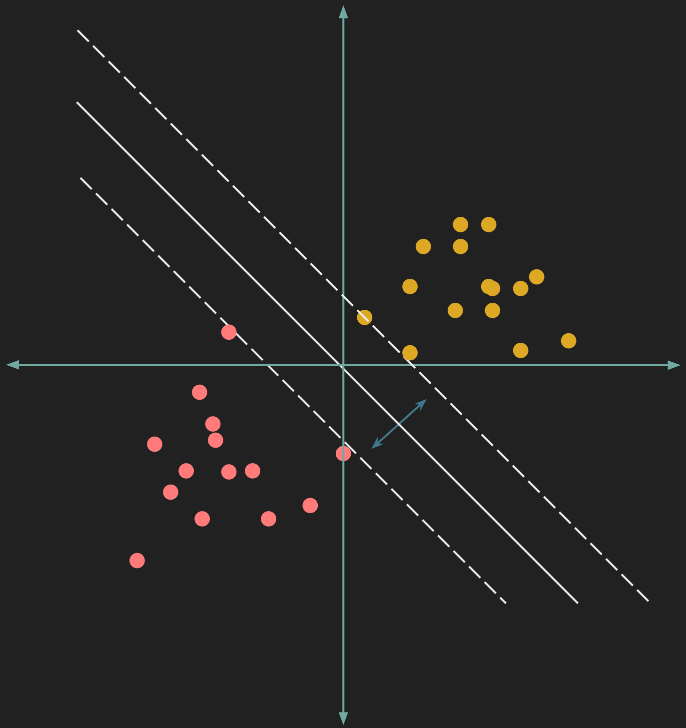
Supervised Learning Example: Support Vector Machine (SVM)

SVMs are a supervised learning example and are used to classify data.

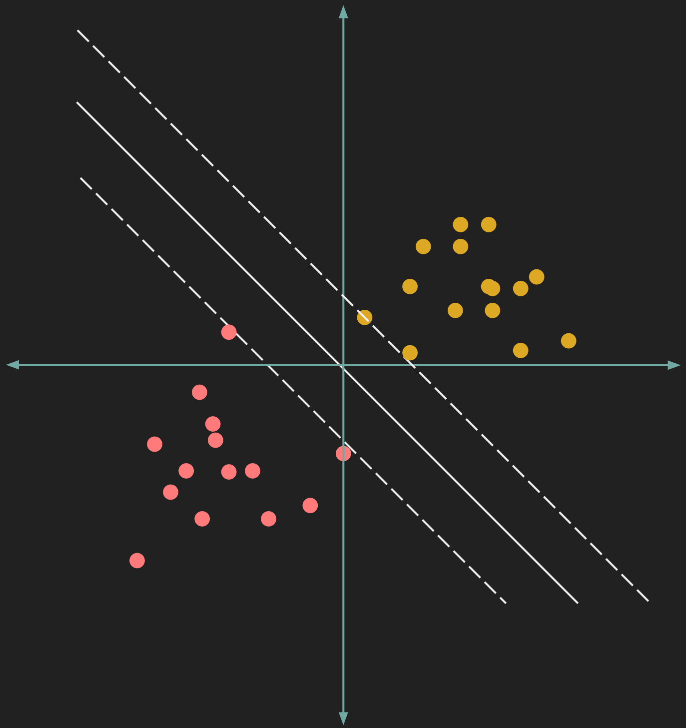


If this is supervised learning, what does that mean about the input data?

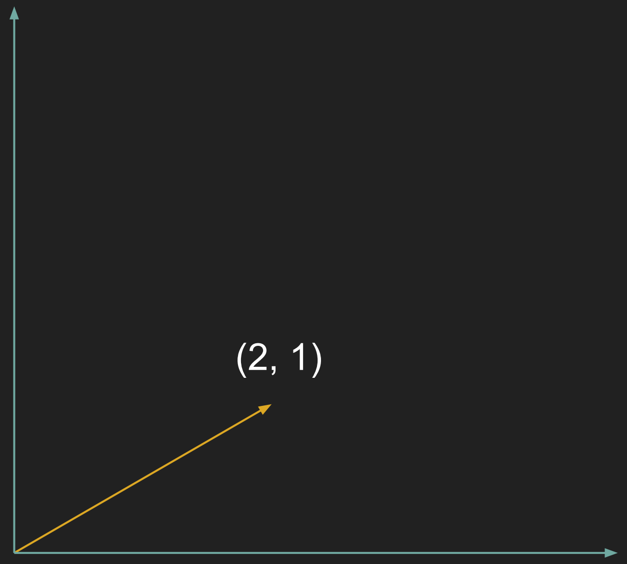
Instead of simply separating the data, it tries to maximize the margin (indicated by the double arrow below) of each of the closest data points to the separating line.



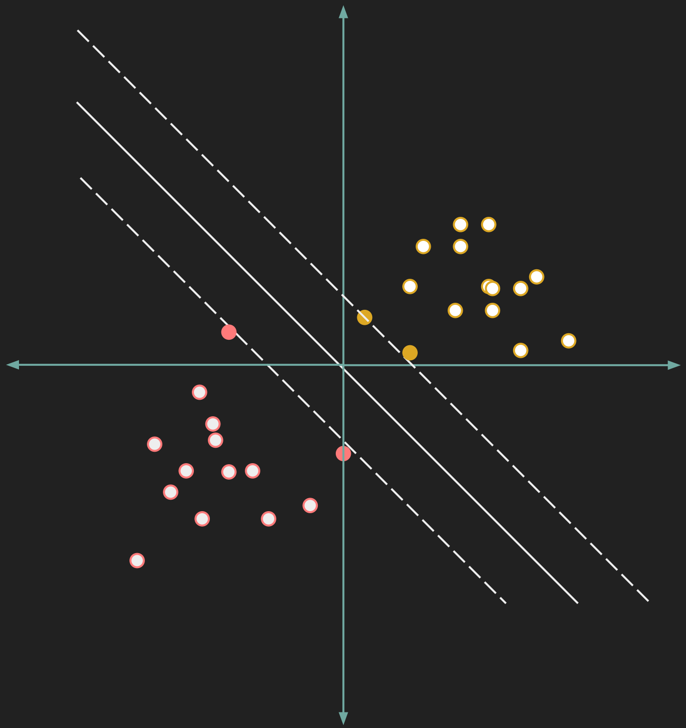
SVMs do this in a similar way to the method used for the parameters of our linear regression example.



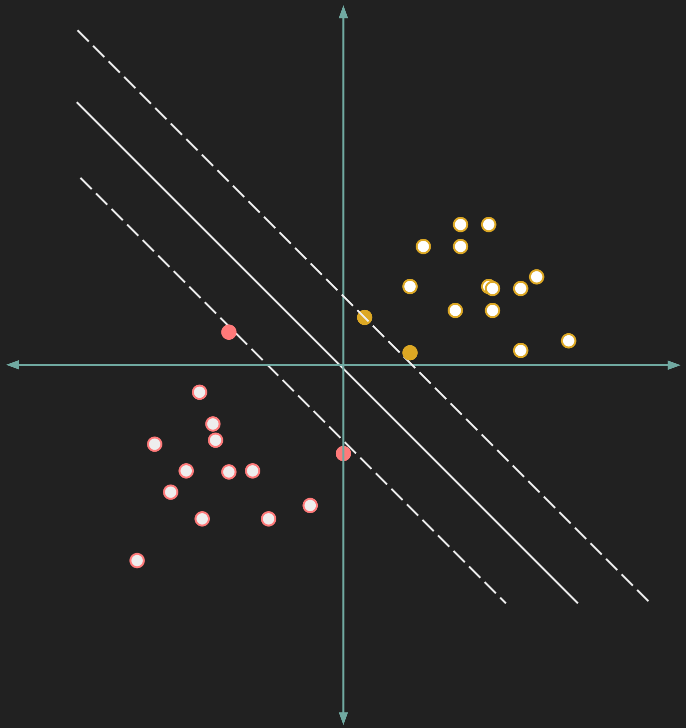
A vector describes a point in space by its direction and magnitude. What is the magnitude of this vector?



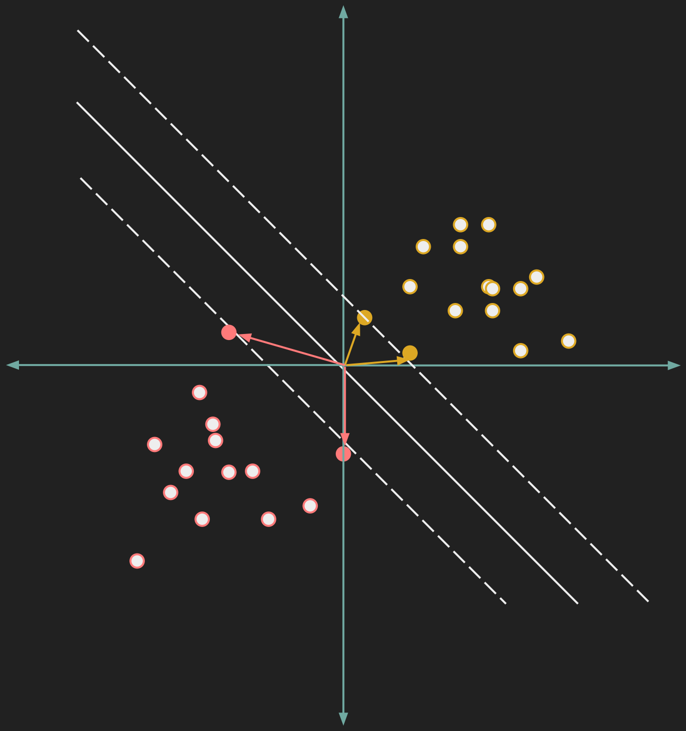
This method is called a SVM because the solution only really depends on the points that define the margin or the support vectors.



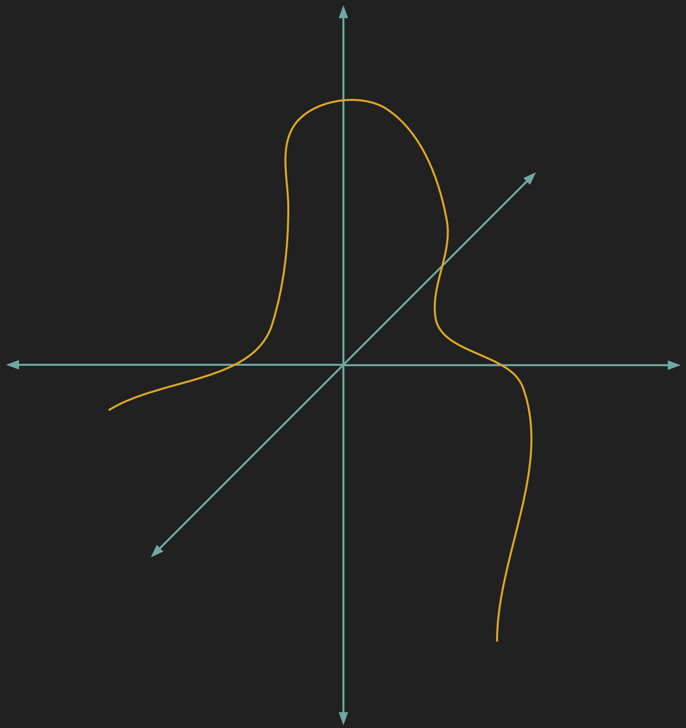
Support vectors are the data points which lie closest to the decision surface.



Here, I have illustrated the actual support vectors, but often just the points are referenced instead.



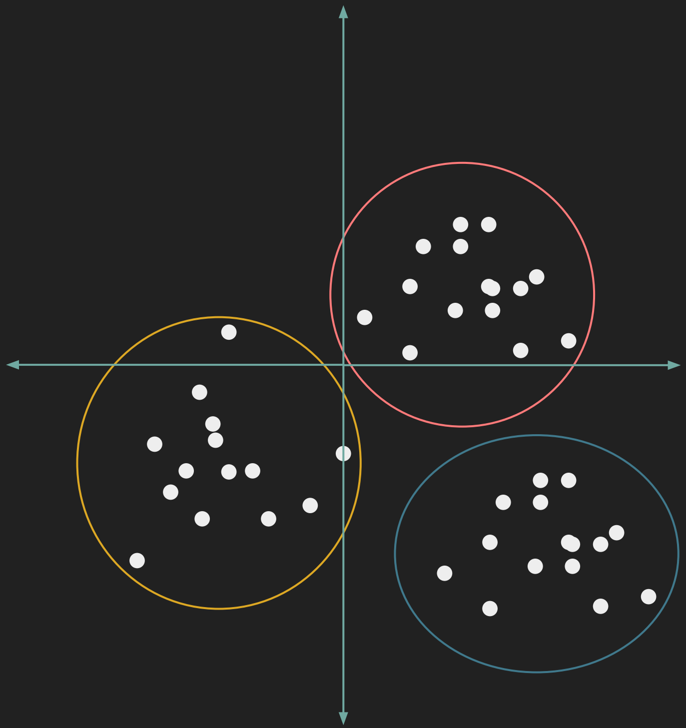
We have illustrated a separating line, but if the dimensions of your data increase, then you will need a separating hyperplane.



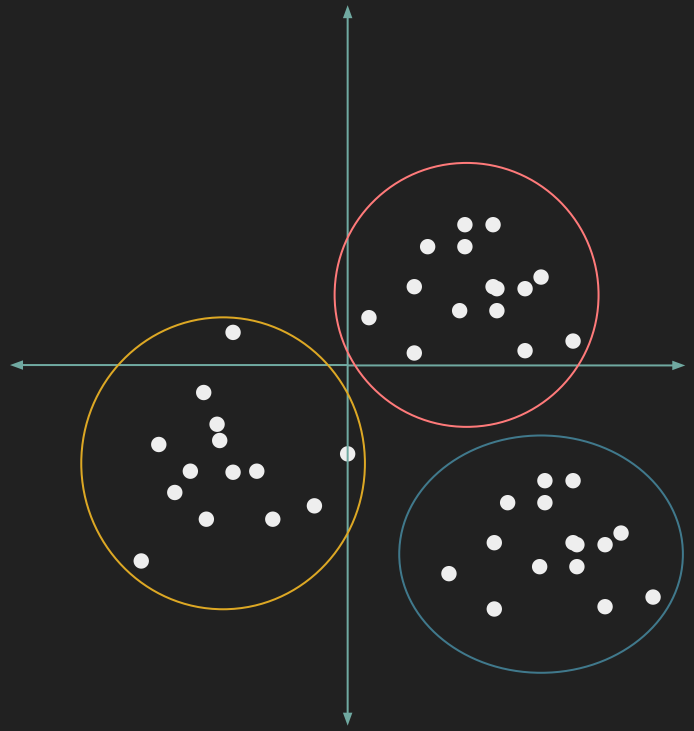
K-means Clustering

Unsupervised Learning: K-means Clustering

K-means clustering is an unsupervised learning example and is also used to classify data.

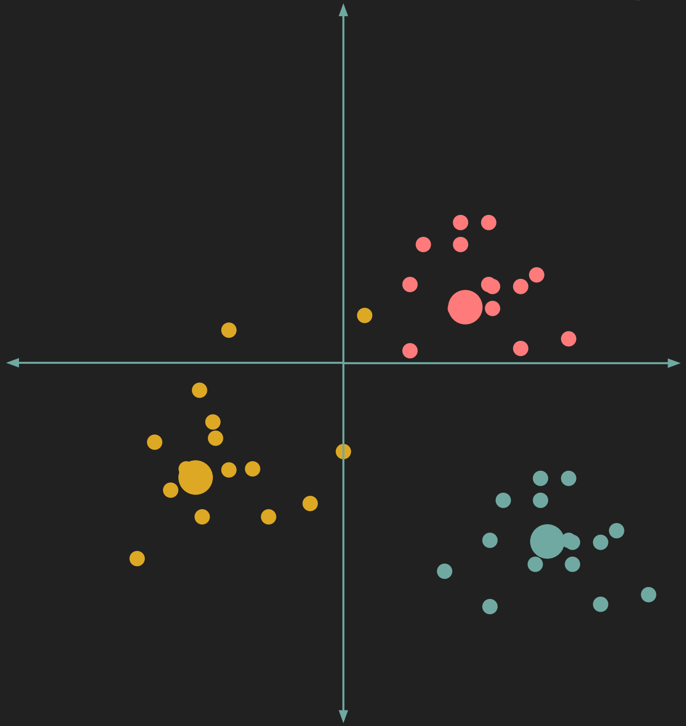


The ultimate goal is to partition the given data into a few groups or clusters.

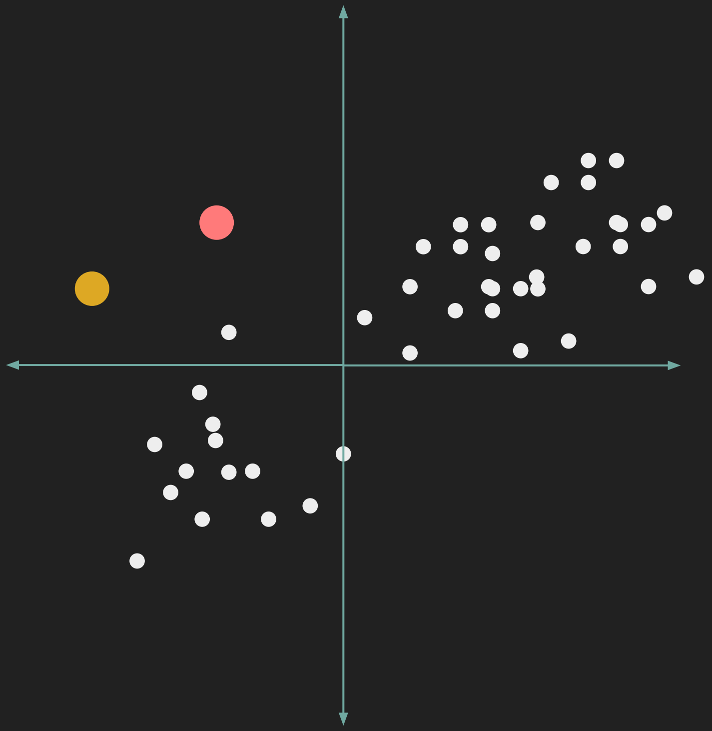


What are some examples of where clustering might be useful?

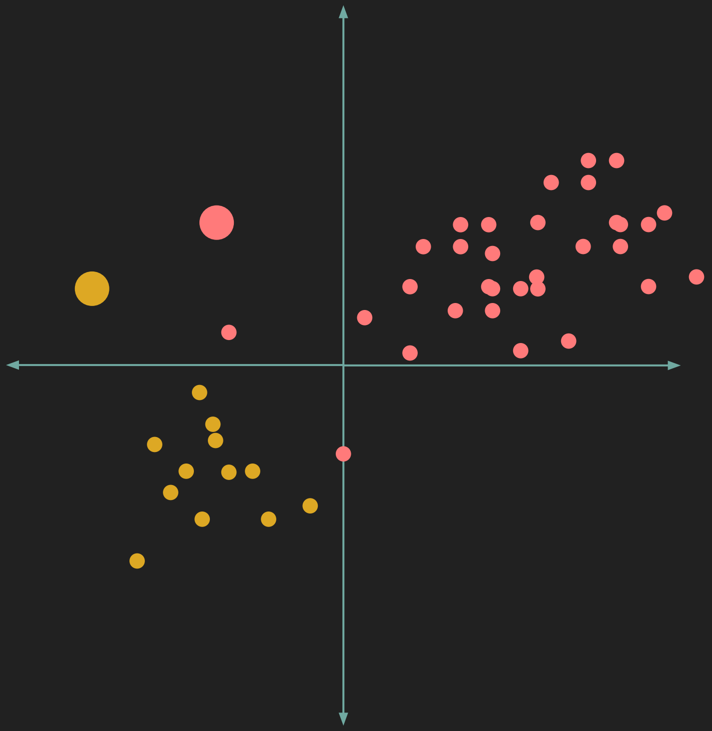
The way that this algorithm learns the appropriate clusters is by placing centroids randomly on the graph, calculating the distance of each point to its closest centroid, recomputing the centroid as the mean of all points and then iterating.



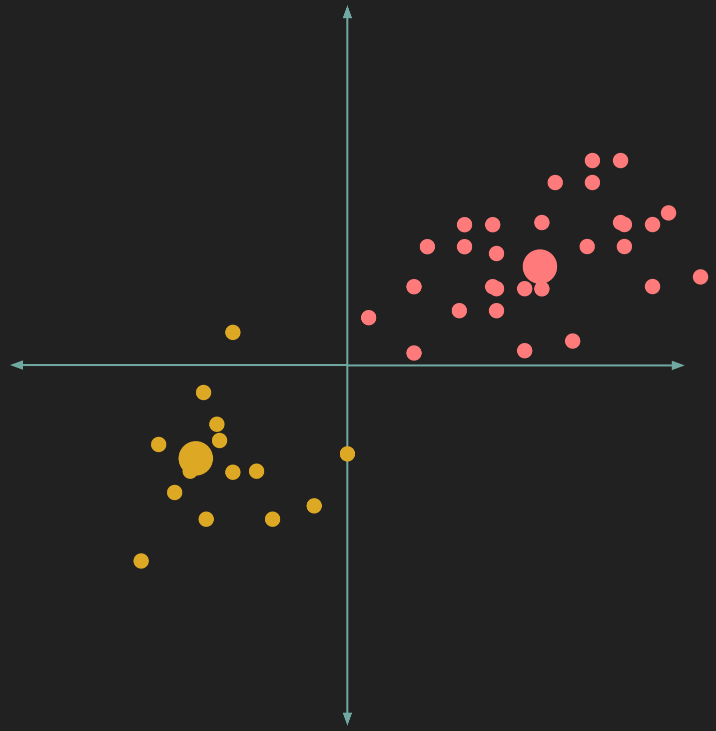
Let us visualize this training procedure.



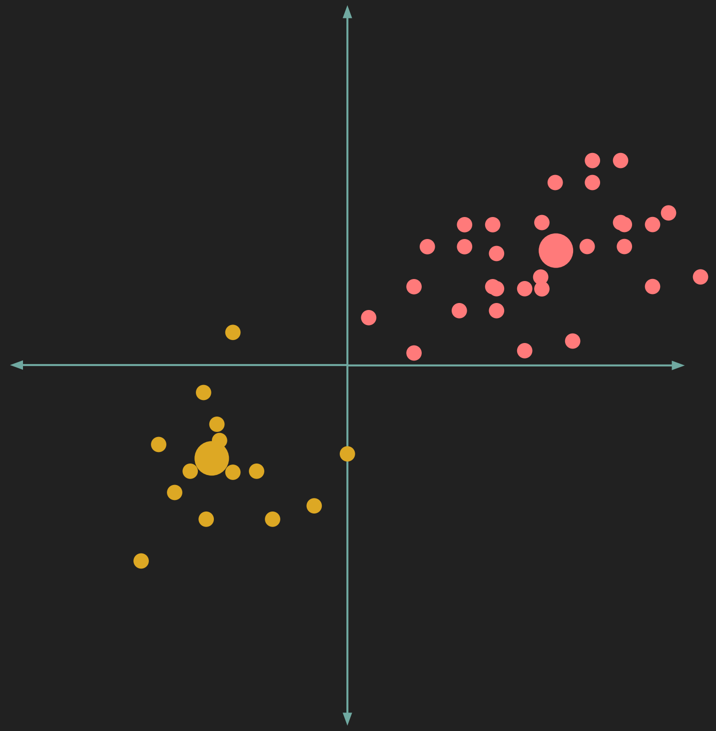
Let us visualize this training procedure.



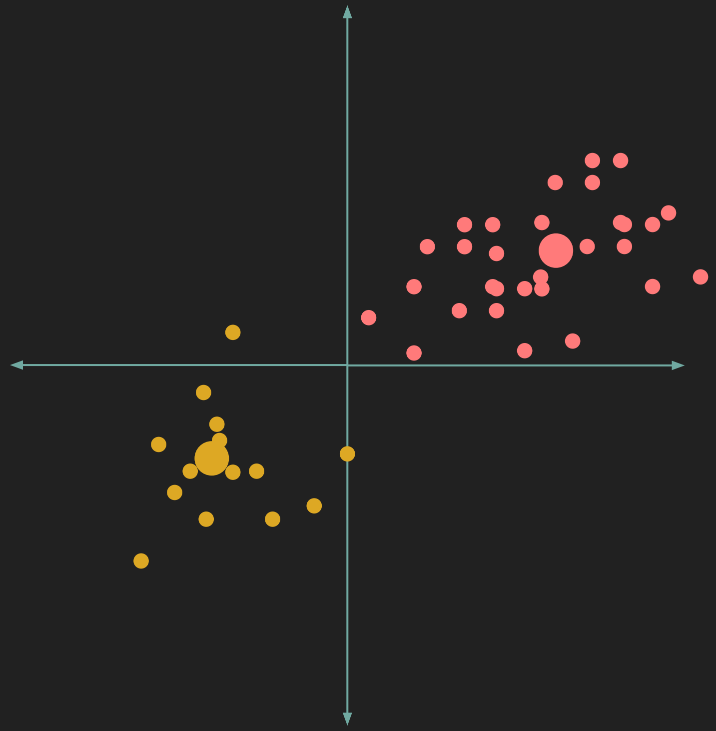
Let us visualize this training procedure.



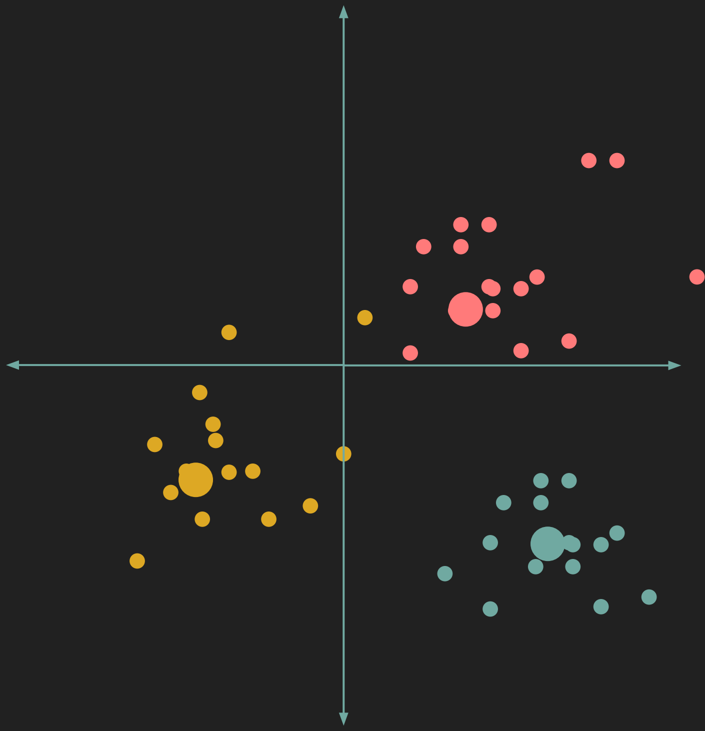
Let us visualize this training procedure.



The number of centroids is a user-prescribed parameter.



The number of centroids is a user-prescribed parameter.



The Experimental Setup

In our experiment we have..

306 sensors

25 different scene images

5 different scene categories

We want to see if we can identify when someone is viewing each scene category, such that we could predict the categories of MEG data from new trials.

If we can do this, this means that the brain is encoding the visual stimuli in a way that is decodable by MEG signals.

If we cannot do this, however, it does NOT mean that the brain is not encoding this information. Why?

Tutorial 1: Support Vector Machine

Tutorial 2: K-means Clustering

Thanks!