Machine Learning

Machine learning is the “field of study that gives computers the ability to learn without being explicitly programmed”.

- Arthur Samuel, Machine Learning Pioneer

Machine Learning

The program improves its performance through *experience*.

Here, experience equates to past information/data.

Types of Machine Learning

Three Main Types of Machine Learning:

1. Supervised Learning
2. Unsupervised Learning
3. Reinforcement Learning

Types of Machine Learning

Supervised Learning:

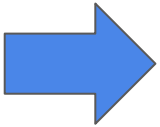
- Labeled data
- Goal: find a function to predict the label/outcome given input data/features.
- Direct feedback

Supervised Learning

How much will this house sell for?

Supervised Learning

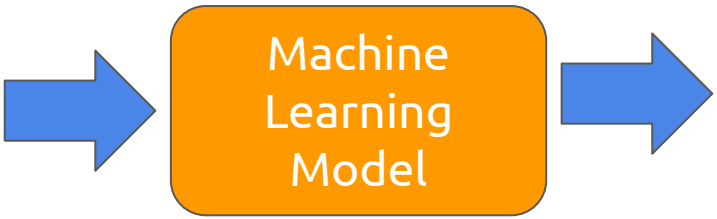
How much will this house sell for?



Machine Learning Model

Supervised Learning

How much will this house sell for?

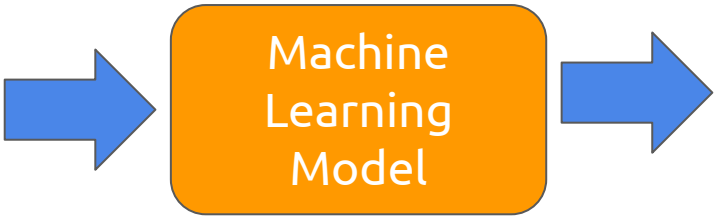


Prediction:
$450,000???

Supervised Learning

You're way off. It's actually $180,000

How much will this house sell for?



Prediction:
$450,000???

Supervised Learning

How much will this house sell for?

Machine Learning Model

Prediction:
$450,000???

I'll try to do better next time.

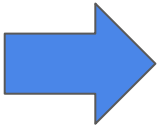
You're way off. It's actually $180,000

Supervised Learning

Try again.
What about this house?

Supervised Learning

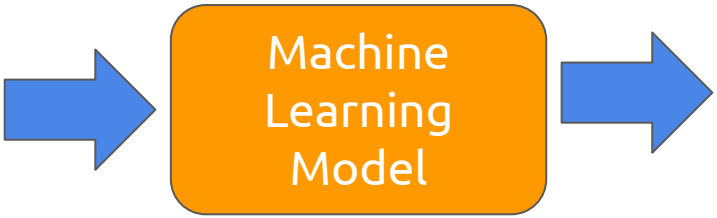
Try again.
What about this house?



Machine Learning Model

Supervised Learning

Try again.
What about this house?

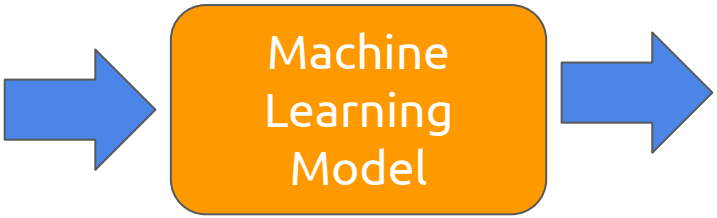


Prediction:
$360,000??

Supervised Learning

You're closer. It's actually $285,000

Try again.
What about this house?



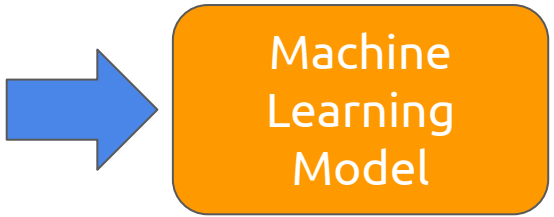
Prediction:
$360,000??

Supervised Learning

After seeing lots of examples, the model gets better at predicting.

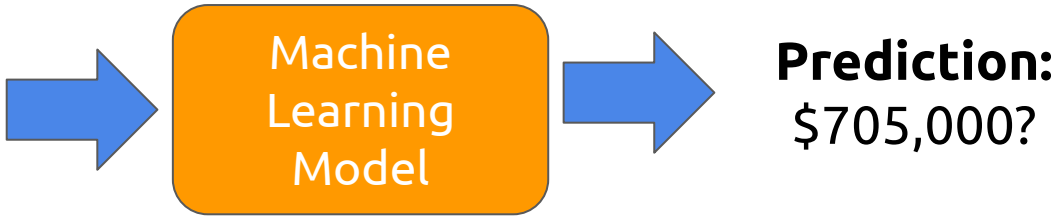
Supervised Learning

After seeing lots of examples, the model gets better at predicting.



Supervised Learning

After seeing lots of examples, the model gets better at predicting.



Supervised Learning

After seeing lots of examples, the model gets better at predicting.

Not bad. It's actually $719,000

Machine Learning Model

Prediction:
$705,000?

Supervised Learning Models

- Linear/Logistic Regression (common statistical models)
- Decision Trees
- Ensemble Methods
- Neural Networks

These will be the focus of the later portions of the course

Types of Machine Learning

Unsupervised Learning:

- Unlabeled data - features but no known outcome
- Goal: find underlying structure in the dataset
- No direct feedback

Types of Unsupervised Learning

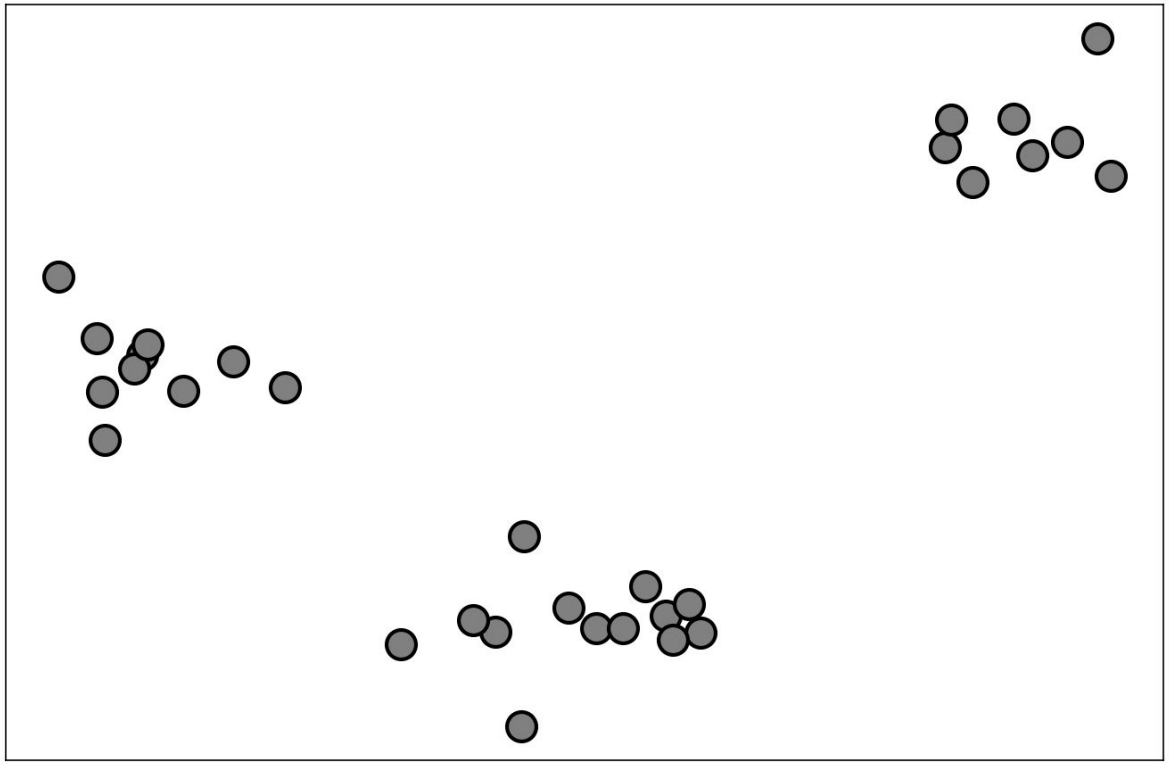
Four main goals:

- **Embedding:** Find a lower-dimensional representation of the data, without losing too much “resolution”
- **Clustering:** discover “clumps” of points (eg. customer segmentation)
- **Density Estimation:** Approximate the probability distribution of the data (think KDE from seaborn).
- Finding good explanations (**hidden causes** or **sources**) of the data (model the data generation process)

Unsupervised Learning Challenges

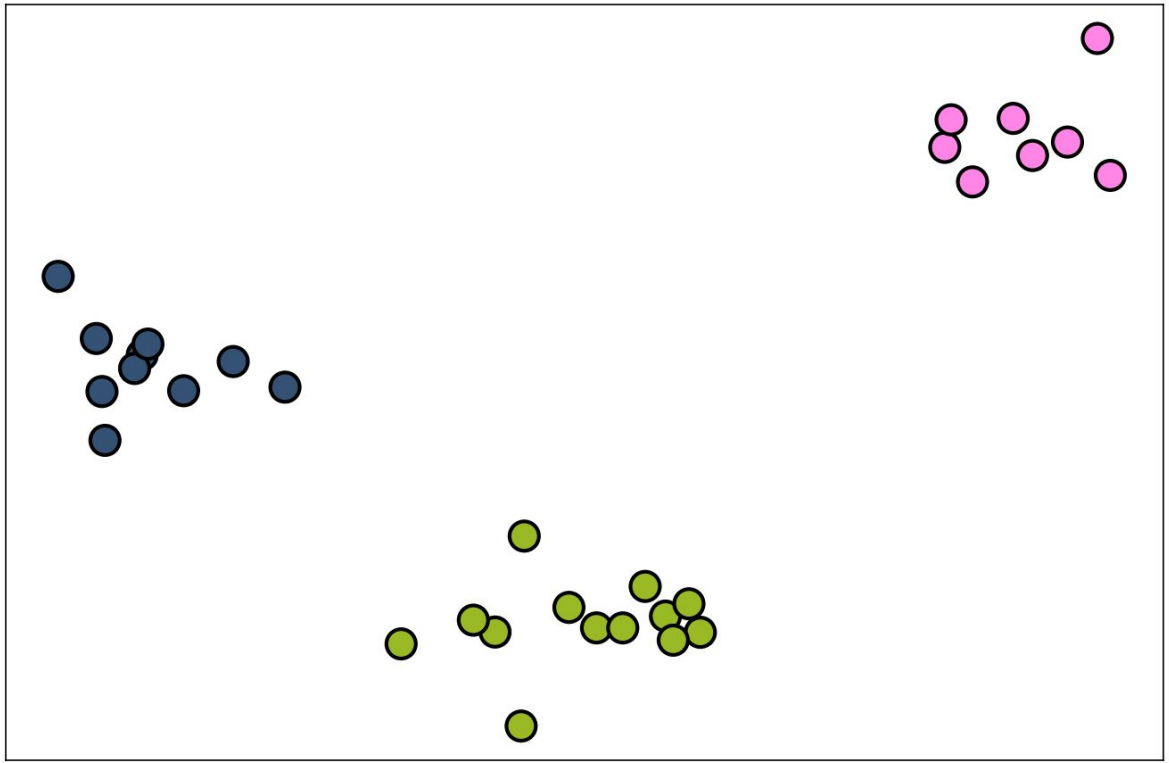
- No simple goal, like with supervised learning.
- No teacher/supervisor to provide “correct answers”.
- Hard to assess the results obtained - no labels, so we can’t compute accuracy or MAE.

Unsupervised Learning



Can be used to uncover groupings or clusters in a dataset.

Unsupervised Learning



Can be used to uncover groupings or clusters in a dataset.

Unsupervised Learning

Can give a lower-dimensional representation of a high-dimensional dataset.

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
0	2.595013	-0.900250	-0.417787	-0.917335	-0.333673	-0.148682	-0.176294	2.574095	-0.085889	-1.572139
1	0.013680	-0.495092	0.447688	1.065596	-1.441302	-0.818622	-0.437475	0.441887	1.393845	-2.387828
2	-0.230319	1.189959	-0.990523	-0.622799	-1.450218	-0.197577	-0.651423	-0.351184	0.142664	1.179185
3	-0.746737	-0.213822	-0.562811	2.582491	-1.555323	-0.443510	-0.284505	-0.762308	-0.387258	0.094279
4	-1.942155	1.025507	0.875429	-1.356284	1.184086	0.357414	0.569523	0.131664	0.719885	0.124905
5	0.475851	0.723243	-0.877883	-1.573716	-0.693388	-0.504333	-1.057207	-0.183671	-0.015274	-0.179363
6	1.129120	0.093178	-0.326481	-0.265502	0.305289	1.726201	-0.033569	0.500447	-1.034352	0.232601
7	2.474474	1.532935	-0.445996	-0.426534	-0.709179	-1.701396	2.255979	0.741835	0.797036	-0.395185
8	-0.086396	-0.254281	0.184538	-0.258440	-1.579722	0.167779	-0.556343	0.725770	-0.273071	-2.273305
9	1.309791	-1.189775	-0.658195	-0.854114	-1.368125	0.299176	-0.040814	0.536397	0.638233	1.533205
10	-0.647555	-0.243093	0.070906	-2.335958	-2.385107	0.782842	0.543407	-0.207614	2.162430	-0.628655
11	1.508285	0.491980	1.465918	1.824885	0.421084	-0.212366	1.389737	2.017906	-1.108551	1.501897
12	1.101790	-0.003095	-0.842642	0.988374	-0.381950	-0.660127	0.679805	-0.399634	0.602226	0.147839
13	0.627606	0.854219	0.412174	-0.197228	0.619822	-0.186393	-1.580918	-0.230607	2.080174	-0.393868

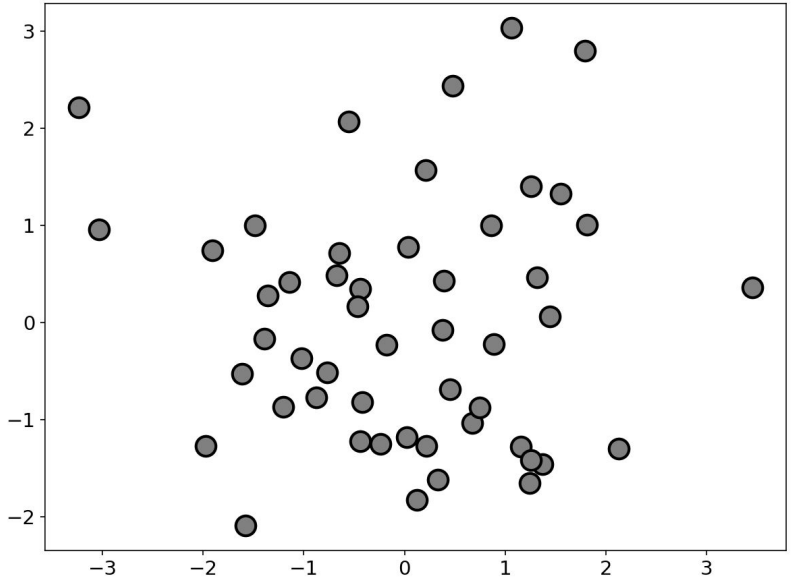
10-Dimensional Dataset

Unsupervised Learning

Can give a lower-dimensional representation of a high-dimensional dataset.

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
0	2.595013	-0.900250	-0.417787	-0.917335	-0.333673	-0.148682	-0.176294	2.574095	-0.085889	-1.572139
1	0.013680	-0.495092	0.447688	1.065596	-1.441302	-0.818622	-0.437475	0.441887	1.393845	-2.387828
2	-0.230319	1.189959	-0.990523	-0.622799	-1.450218	-0.197577	-0.651423	-0.351184	0.142664	1.179185
3	-0.746737	-0.213822	-0.562811	2.582491	-1.555323	-0.443510	-0.284505	-0.762308	-0.387258	0.094279
4	-1.942155	1.025507	0.875429	-1.356284	1.184086	0.357414	0.569523	0.131664	0.719885	0.124905
5	0.475851	0.723243	-0.877883	-1.573716	-0.693388	-0.504333	-1.057207	-0.183671	-0.015274	-0.179363
6	1.129120	0.093178	-0.326481	-0.265502	0.305289	1.726201	-0.033569	0.500447	-1.034352	0.232601
7	2.474474	1.532935	-0.445996	-0.426534	-0.709179	-1.701396	2.255979	0.741835	0.797036	-0.395185
8	-0.086396	-0.254281	0.184538	-0.258440	-1.579722	0.167779	-0.556343	0.725770	-0.273071	-2.273305
9	1.309791	-1.189775	-0.658195	-0.854114	-1.368125	0.299176	-0.040814	0.536397	0.638233	1.533205
10	-0.647555	-0.243093	0.070906	-2.335958	-2.385107	0.782842	0.543407	-0.207614	2.162430	-0.628655
11	1.508285	0.491980	1.465918	1.824885	0.421084	-0.212366	1.389737	2.017906	-1.108551	1.501897
12	1.101790	-0.003095	-0.842642	0.988374	-0.381950	-0.660127	0.679805	-0.399634	0.602226	0.147839
13	0.627606	0.854219	0.412174	-0.197228	0.619822	-0.186393	-1.580918	-0.230607	2.080174	-0.393868

10-Dimensional Dataset

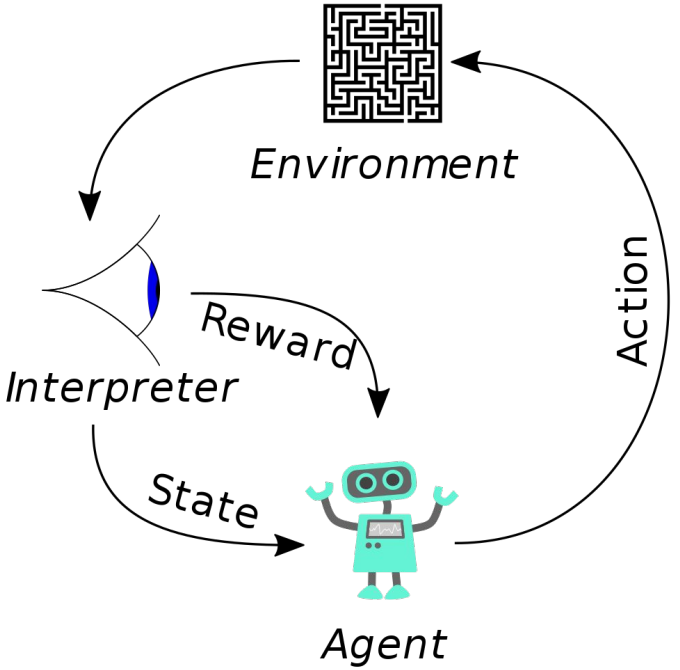


2-Dimensional Representation

Types of Machine Learning

Reinforcement Learning:

An **agent** interacting with an **environment** by taking **actions** with the goal of maximizing a cumulative **reward**.



https://en.wikipedia.org/wiki/Reinforcement_learning#/media/File:Reinforcement_learning_diagram.svg