Welcome to the Course!

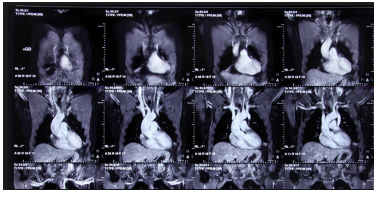
Lecture 1 - DAMLF | ML1

Machine Learning & Business

Transforming internet businesses

- web search
- advertising
- recommendation

Machine Learning & Business



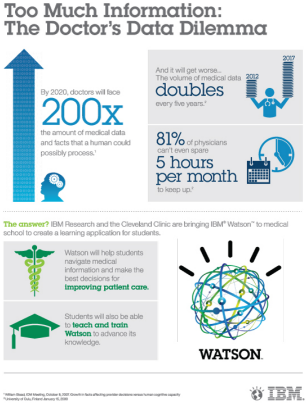
Transforming internet businesses

- web search
- advertising
- recommendation

Enabling new products

- better health care
- precision agriculture / satellite farming
- self-driving cars
- *and more* ...

Machine Learning & Business



Transforming internet businesses

- web search
- advertising
- recommendation

Enabling new products

- better health care
- precision agriculture / satellite farming
- self-driving cars
- *and more* ...

Machine Learning & Business

Transforming internet businesses

- web search
- advertising
- recommendation

Enabling new products

- better health care
- precision agriculture / satellite farming
- self-driving cars
- *and more* ...

Machine Learning & Business

Larry Page's air taxis are already flying above New Zealand

A new all-electric autonomous aircraft built by the startup Kitty Hawk has been in testing for months.

Transforming internet businesses

- web search
- advertising
- recommendation

Enabling new products

- better health care
- precision agriculture / satellite farming
- self-driving cars
- *and more* ...

Machine Learning & Finance

CNBC make it

WORK

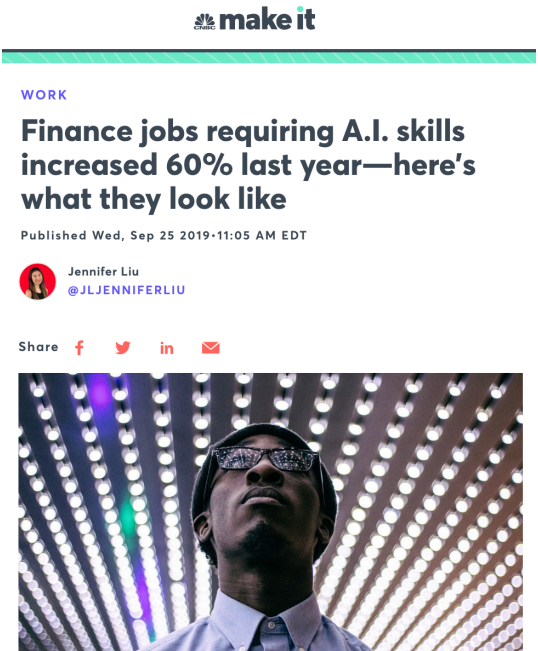
Finance jobs requiring A.I. skills increased 60% last year—here's what they look like

Published Wed, Sep 25 2019•11:05 AM EDT

Jennifer Liu
@JLJENNIFERLIU

Share

Machine Learning & Finance



make it

WORK

Finance jobs requiring A.I. skills increased 60% last year—here's what they look like

Published Wed, Sep 25 2019·11:05 AM EDT

Jennifer Liu
@JLJENNIFERLIU

Share

Trading: Algo & High Frequency

Robo-advisors
- Chatbots
- Portfolio Management

Risk Management
- Loan Credit Scoring
- Insurance Underwriting

Anomaly Detection
- Fraud Detection
- Money Laundering Prevention
- Predicting / Resolving Failed Trades

Document Analysis
- FinReg
- Automating Workflows

Exogenous Sources Data
- Macro Economic
- Sentiment / Social Networks
- Imaging / Video
- IoT

What is a Learning Algorithm?

"A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E"

–Tom Mitchell

Ingredients:

Task *to perform*

Type of **Experience**

Performance *measure*

What is a Learning Algorithm?

"A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E"

–Tom Mitchell

Ingredients:

Task *to perform*

Type of **Experience**

Performance *measure*

Tasks

- Regression
- Classification
- Prediction
- Pattern Recognition / Discovery

Ingredients:

Task *to perform*

Type of **Experience**

Performance *measure*

Types of Experiences

- Supervised Learning
- Unsupervised Learning
- Reinforcement

Ingredients:

Task *to perform*

Type of **Experience**

Performance *measure*

Performance

- A Numerical Measure
- Mathematical Optimization

What is Machine Learning?

Areas of AI

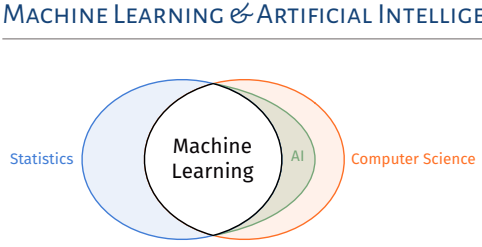
- Knowledge Representation & Reasoning
- Planning
- Cognitive Modeling
- Agent-based Modeling
- Robotics
- Vision / Pattern Recognition
- Natural Language Processing
- Foundations of AI

Machine Learning & Artificial Intelligence

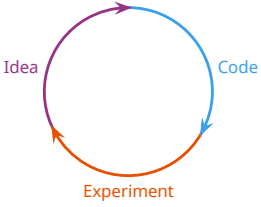
What is Machine Learning?

Areas of AI

- Knowledge Representation & Reasoning
- Planning
- Cognitive Modeling
- Agent-based Modeling
- Robotics
- Vision / Pattern Recognition
- Natural Language Processing
- Foundations of AI



This Course



Builds on What you Know

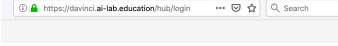
- Rudimentary Statistics
- Some Ideas from Calculus
- Familiarity with Python

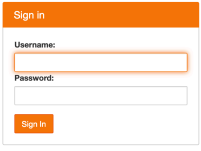
This Course

Prof. Dr. Gregory Wheeler

TA: Mr. Chukwuma (Chuks) Dim

This Course - AI Lab Servers

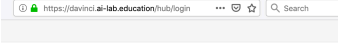


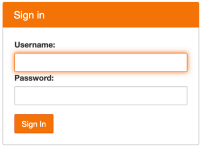


1. Select your course

- Machine Learning I
  https://raphael.ai-lab.education/
- Data Analytics and Machine Learning in Finance
  https://delacroix.ai-lab.education/

This Course - AI Lab Servers





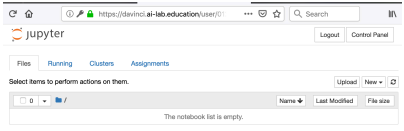
1. Select your course

- Machine Learning I
  https://raphael.ai-lab.education/
- Data Analytics and Machine Learning in Finance
  https://delacroix.ai-lab.education/

2. Sign In

- Machine Learning I
  Your user name is your student ID number
- Data Analytics and Machine Learning in Finance
  Your user name is the letter 'd' or 'm' and your student ID number

This Course - AI Lab Servers



1. Select your course

- Machine Learning I
  https://raphael.ai-lab.education/
- Data Analytics and Machine Learning in Finance
  https://delacroix.ai-lab.education/

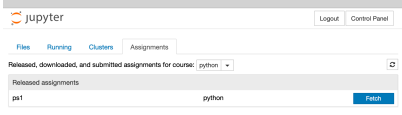
2. Sign In

- Machine Learning I
  Your user name is your student ID number
- Data Analytics and Machine Learning in Finance
  Your user name is the letter ‘d‘ or ‘m‘ and your student ID number

2. Fetch Assignments

- Once in, click on **Assignments**

This Course - AI Lab Servers



1. Select your course

- Machine Learning I
  https://raphael.ai-lab.education/
- Data Analytics and Machine Learning in Finance
  https://delacroix.ai-lab.education/

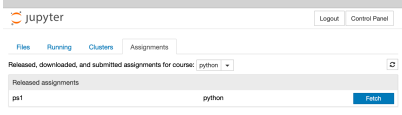
2. Sign In

- Machine Learning I
  Your user name is your student ID number
- Data Analytics and Machine Learning in Finance
  Your user name is the letter 'd' or 'm' and your student ID number

2. Fetch Assignments

- Once in, click on **Assignments**
- Then, **Fetch** available assignments

This Course - AI Lab Servers



1. Select your course

- Machine Learning I
  https://raphael.ai-lab.education/
- Data Analytics and Machine Learning in Finance
  https://delacroix.ai-lab.education/

2. Sign In

- Machine Learning I
  Your user name is your student ID number
- Data Analytics and Machine Learning in Finance
  Your user name is the letter 'd' or 'm' and your student ID number

2. Fetch Assignments

- Once in, click on **Assignments**
- Then, **Fetch** available assignments
- Or watch **this video**

This Course - Deadlines

Assignment	Topic	m d due dates	points
ps1	Linear Regression & Gradient Descent	m 23.feb.2020 d 19.feb.2020	14
ps2	Binary & Multiclass Classification	m 01.mar.2020 d 26.feb.2020	14
ps3	Decision Trees & Random Forests	m 08.mar.2019 d 04.mar.2020	14
ps4	Neural Networks	m 19.mar.2020 d 15.mar.2020	14
ps5	K-means Clustering	m 25.mar.2020 d 28.mar.2020	14
	Final Exam	xx.apr.2020	50

This Course - Policies

Collaboration

- No more than 4 people in a group
- Discuss but do not copy
- Always list your collaborators.

Late Assignments

- You have a 30 hour "late budget" for the course
- You may be 30 hours accumulatively late on assignments without penalty.
- After 30 hours, each late assignment receives 0 points.

Course Syllabus

- Available on Canvas.

This Course - Materials

Lecture Notes

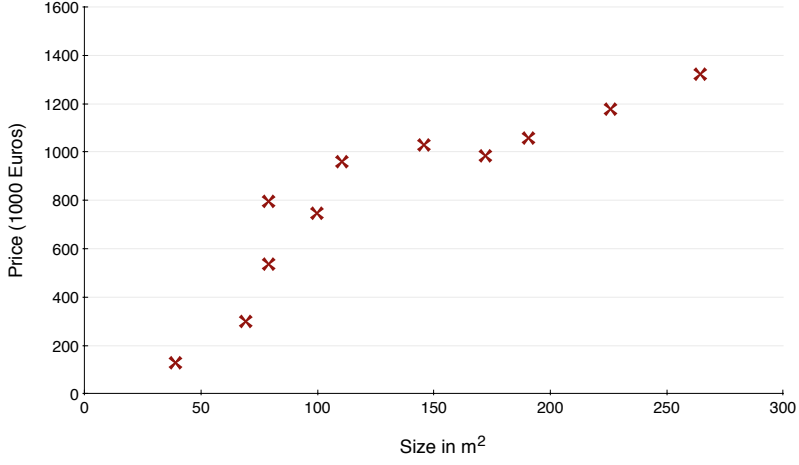
Wheeler, G. (2020)
Lecture Notes for Machine Learning: An Introduction to Theory and Applications
Available on Canvas

Online Book

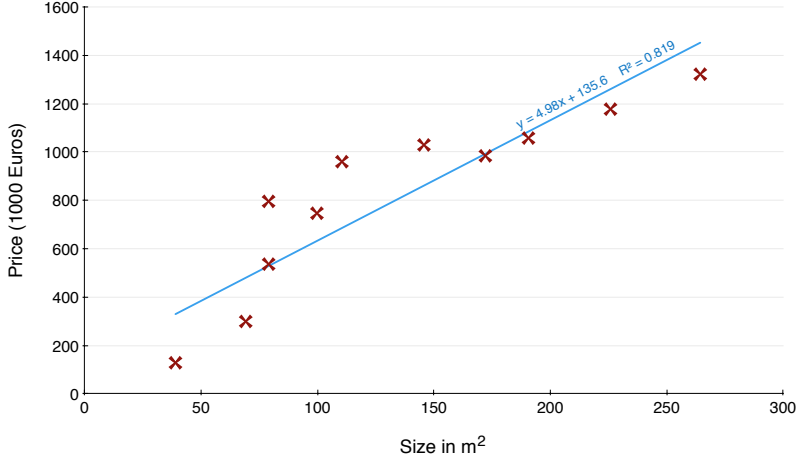
Nelson, M. (2019)
Neural Networks and Deep Learning
neuralnetworksanddeeplearning.com

Let’s begin!

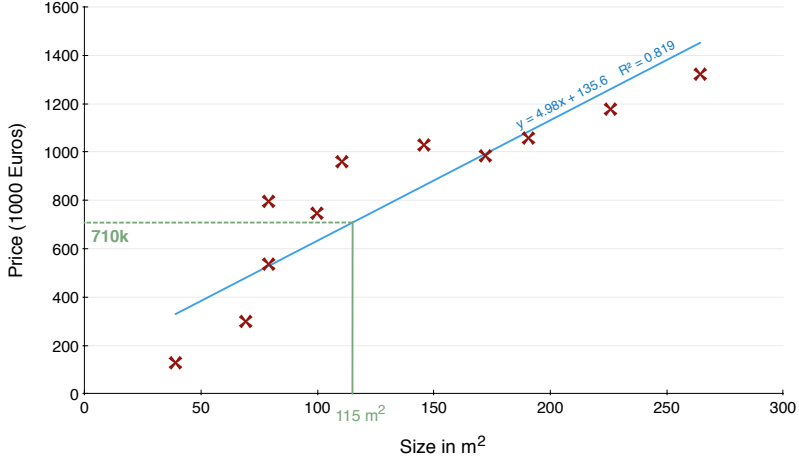
House Prices



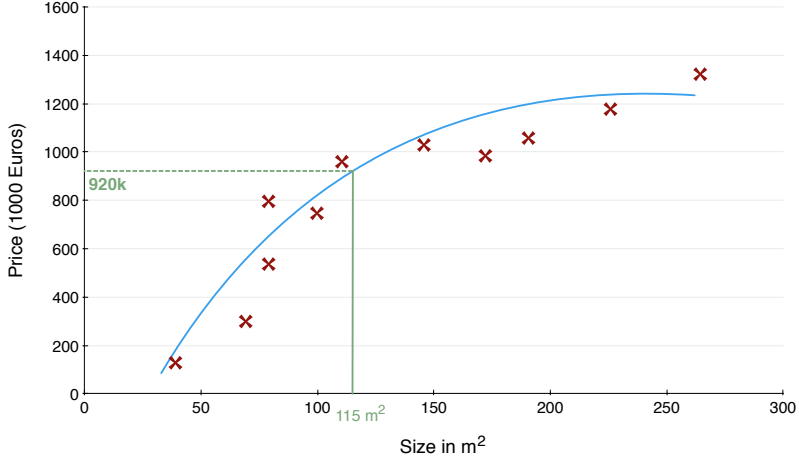
House Prices



House Prices



House Prices



Regression:

- y: *response*
- x: *feature*
- θ_0, θ_1: *parameters*
- ϵ: *noise*

Regression Problem

Data Generating Process:

$$y = \theta_0 + \theta_1 x + \epsilon$$

Regression:

- y: *response*
- x: *feature*
- θ_0, θ_1: *parameters*
- ϵ: *noise*

Regression Problem

Data Generating Process:

$$y = \underbrace{\theta_0 + \theta_1 x}_{\text{structural component}} + \underbrace{\epsilon}_{\text{error component}}$$

Regression:

- y: *response*
- x: *feature*
- θ_0, θ_1: *parameters*
- ϵ: *noise*

Regression Problem

Data Generating Process:

$$y = \underbrace{\theta_0 + \theta_1 x}_{\text{structural component}} + \underbrace{\epsilon}_{\text{error component}}$$

Linear model:

$$\hat{y} = \hat{\theta}_0 + \hat{\theta}_1 x$$

Regression:

- y: *response*
- x: *feature*
- θ_0, θ_1: *parameters*
- ϵ: *noise*

Regression Problem

Data Generating Process:

$$y = \underbrace{\theta_0 + \theta_1 x}_{\text{structural component}} + \underbrace{\epsilon}_{\text{error component}}$$

Linear model:

$$\begin{aligned}\hat{y} &= \hat{\theta}_0 + \hat{\theta}_1 x \\ &= h(x; \hat{\boldsymbol{\theta}})\end{aligned}$$

Regression:

- y: *response*
- x: *feature*
- θ_0, θ_1: *parameters*
- ϵ: *noise*

Regression Problem

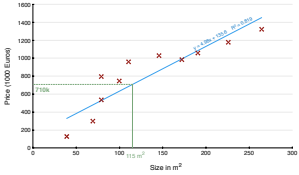
Data Generating Process:

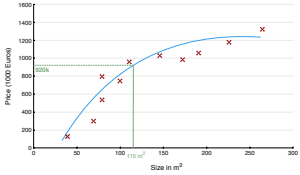
$$y = \underbrace{\theta_0 + \theta_1 x}_{\text{structural component}} + \underbrace{\epsilon}_{\text{error component}}$$

Linear model:

$$\begin{aligned} \hat{y} &= \hat{\theta}_0 + \hat{\theta}_1 x \\ &= \underbrace{h(x; \hat{\boldsymbol{\theta}})}_{\text{parametric model}} \end{aligned}$$

Viewed as a learning problem





- Could fit a **straight line** through the data.
- But, this isn't the only learning algorithm you could use.
- We might think that a **quadratic function** (a second-order polynomial) is a better fit.
- We will talk later about how to choose which function to fit.
- Predicting a selling price is an example of a **supervised learning problem**.

Supervised learning

Q: *What is a supervised learning problem?*

A: A labeled dataset that includes the "right" answers.

- In the house price dataset, for *every* house's size (x) in m^2, we are told the price y in Euros.

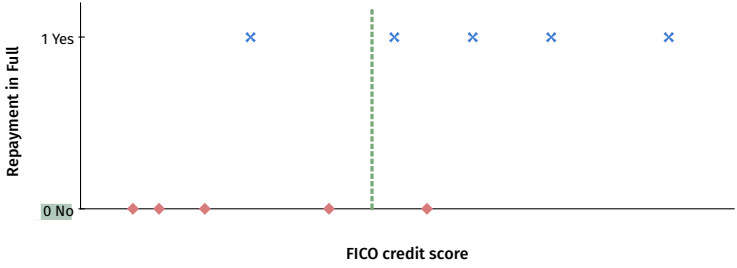
$$(40.1, 125.8)$$
$$(79.0, 536.5)$$
$$(172.1, 984.3)$$
$$\vdots$$

- This is a **regression problem**: the aim is to predict a continuous valued output (selling price) for a house that has *not* been sold yet.

Supervised Learning

- Suppose a loan applicant has a FICO score of 570.
- The ML problem is to estimate what the probability is that a loan applicant with *this* FICO score will replay the loan in full.
- This is a **classification problem**.

Supervised Learning



- Suppose a loan applicant has a FICO score of 570.
- The ML problem is to estimate what the probability is that a loan applicant with *this* FICO score will replay the loan in full.
- This is a **classification problem**.

Classification Problems

Q: *What is a classification problem?*

A: A type of supervised learning problem that aims to predict a discrete value.

- In the loan repayment example, the categories were '0' (default) and '1' (no default).
- But, in general it can be any number of discrete values.

Feature dimension

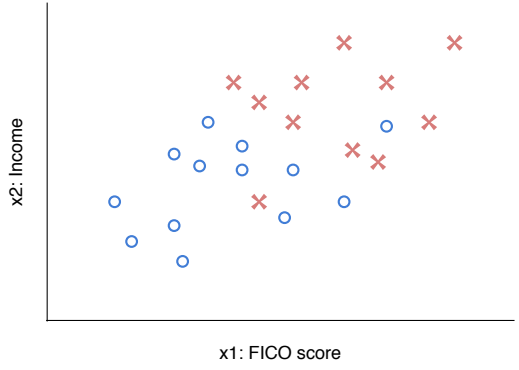
A vector of 1's and 0's is called a **feature dimension**.

For example, *FICO Score* is a feature dimension of customers of some credit product.

Multiple features

Feature dimensions allow us to consider classification problems with **two or more features**.

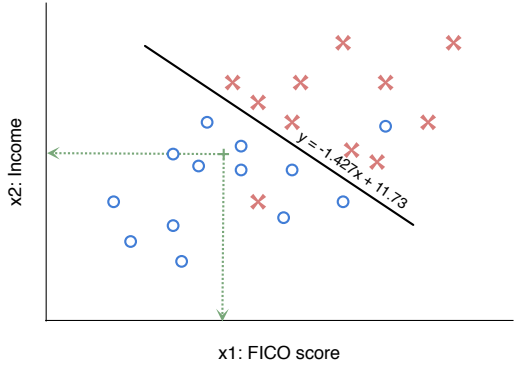
Suppose we know the **income** and **FICO** score for each customer.



Multiple features

Feature dimensions allow us to consider classification problems with **two or more features**.

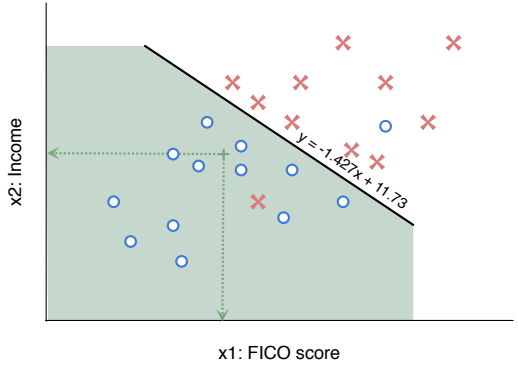
Suppose we know the **income** and **FICO** score for each customer.



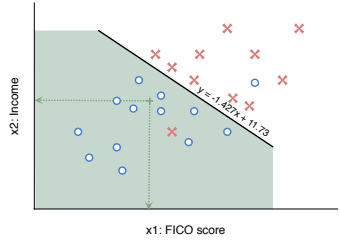
Multiple features

Feature dimensions allow us to consider classification problems with **two or more features**.

Suppose we know the **income** and **FICO** score for each customer.



Classification with two features



An algorithm that fits a straight line to data for predicting full repayment with **two features**, x_1 = FICO score, and x_2 = income, would predict this customer (**+**) is likely to default.

Classification with many features

In a more realistic example, a company may have *many* features:

- Age
- Income
- Gender
- Credit Scores
- Postal Code
- Education
- Marital Status
- Religion
- Social Network Psychological Profile
- Sexual Orientation
- Political Affiliation
- Health
- *et cetera*

Classification with many features

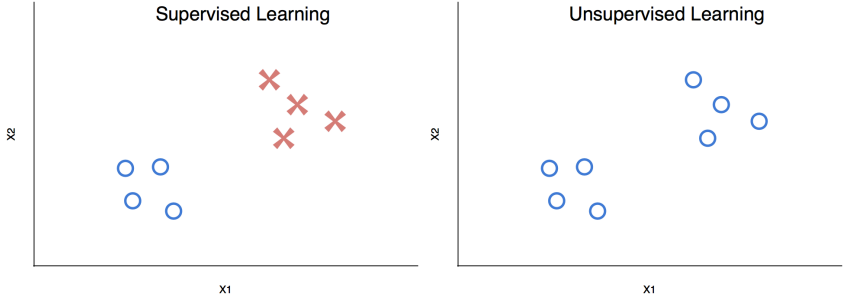
In a more realistic example, a company may have *many* features:

- Age
- Income
- Gender
- Credit Scores
- Postal Code
- Education
- Marital Status
- Religion
- Social Network Psychological Profile
- Sexual Orientation
- Political Affiliation
- Health
- *et cetera*

Some of these features are illegal to use in some countries, but not illegal in all.

Unsupervised Learning

Supervised learning problems involve datasets where each training example is labeled.
Unsupervised learning problems involve datasets which are unlabeled.



An unsupervised ML problem is this: given some data, can you find some structure?

Examples of Unsupervised Learning

News aggregators (Google News).
Looks at thousands of news stories, and **clusters** similar stories.

Examples of Unsupervised Learning

News aggregators (Google News).
Looks at thousands of news stories, and **clusters** similar stories.

Facial Recognition
Look at unlabeled images/video and **clusters** pictures together of people who "closely" resemble one another.

Recap

Supervised Learning:

- Experience a dataset containing **features** $x_1, x_2, \ldots, x_n$.
- Each example is associated with a **label**.
- The goal is to **predict** the unobserved value of an object given observed values for the x_i's.
- Two types of supervised learning problems:
  - **Regression problems**
    predict continuous (scalar) values for y
  - **Classification problems**
    predict discrete values for y

Recap

Unsupervised Learning:

- Experience a dataset containing **features** $x_1, x_2, \ldots, x_n$.
- The goal is to **learn** useful properties of the structure of the x_i's.
- Some types of supervised learning problems:
  - Clustering
  - Anomoly Detection
  - Principal Component Analysis (PCA)

Another Approach

Reinforcement Learning:

- Algorithms **interact** with the environment.
- Creates a feedback loop between the learning system and its experiences.
- The goal is to **adapt** to a decision environment.
  - Agent-based modeling
  - Game AI
  - Evolutionary game theory

Beyond the scope of this course.