

CSCI 111 Web Programming and Problem Solving

Section 1

PART III Artificial Intelligence

Weeks [12 - 15]

Week-12-lecture-2: Machine Learning

Instructor: Dr. Talgat Manglayev

CONTENT

- Introduction
- Supervised Learning
 - Regression
 - Classification
- Unsupervised Learning
 - Clustering
 - Dimensionality reduction
 - Generative ML
- Reinforcement Learning

Introduction



STORE DATA

PROCESS DATA

MAKE DECISION



STORE DATA

PROCESS DATA

+ Machine Learning

MAKE DECISION

INTRODUCTION

- **Predicting house prices based on the house's size, number of rooms, and location**
- **Predicting today's stock market prices based on yesterday's prices and other factors of the market**
- **Detecting spam and non-spam emails based on the words in the e-mail and the sender**
- **Recognizing images as faces or animals, based on the pixels in the image**
- **Processing long text documents and outputting a summary**
- **Recommending videos or movies to a user (e.g., on YouTube or Netflix)**
- **Building chatbots that interact with humans and answer questions**
- **Training self-driving cars to navigate a city by themselves**
- **Diagnosing patients as sick or healthy**
- **Segmenting the market into similar groups based on location, acquisitive power, and interests**
- **Playing games like chess or Go**

INTRODUCTION

- Supervised learning
 - Regression
 - Classification
- Unsupervised learning
 - Clustering
 - Dimensionality reduction
 - Generative ML
- Reinforcement learning

INTRODUCTION. Terms.

Data - any information we use to build **model**

Feature - properties of the data

Label - a particular feature to be predicted by the model

Prediction - the “*guess*” of the model

INTRODUCTION. Terms.

Data - any information we use to build model

Labeled data - has useful information

Unlabeled data - has no useful information relevant to the model

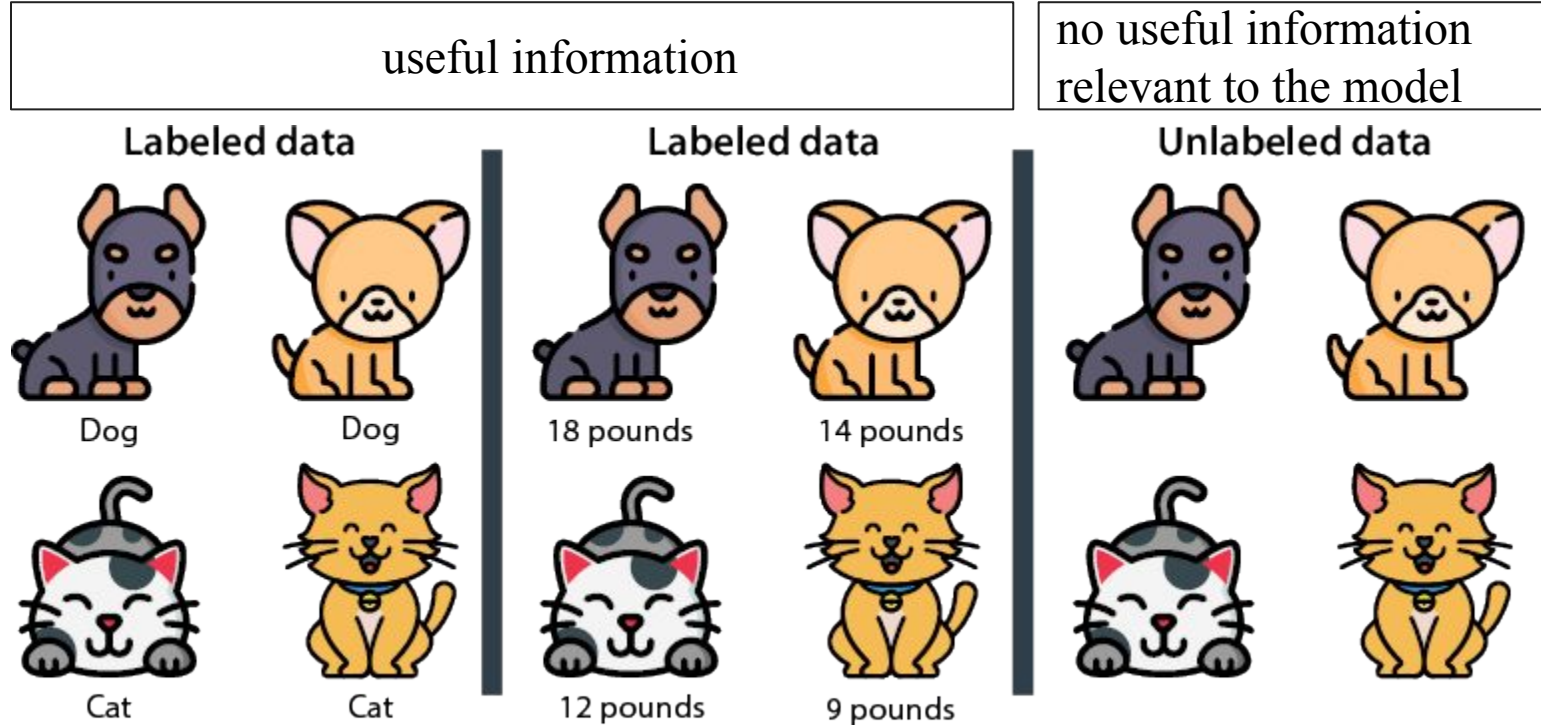
Feature - properties of the data

Label - a particular feature to be predicted by the model

Prediction - the “*guess*” of the model

INTRODUCTION. Terms.

Data - any information we use to build model



Supervised Learning - predict using labeled data

Classification model use
categorical data

Regression model use
numerical data

Labeled data



Dog



Dog



Cat



Cat

Labeled data



18 pounds



14 pounds



12 pounds



9 pounds

Unlabeled data



Supervised Learning Regression

- **Housing Prices** - predict price of a new house, based on prices and features of existing houses
- **Stock Market** - predict price of a certain stock
- **Medicine** - predict expected recovery time
- **Sales** - predict expected amount of money customer will spend
- **Video Recommendation** - predicted the expected amount of time a user will watch

Examples

- **Decision tree**
- **Random Forests**
- **AdaBoost**
- **Gradient Boost Trees**
- **XGBoost**
- **etc.**

Methods

Supervised Learning Classification Examples

- Predict the type of pet (cat, dog)
- Predict if email is spam (spam or not)
- Image recognition MNIST
- Image recognition CIFAR
- Website traffic (whether user will follow the link)
- Social Media (whether user will be friend)
- Video Recommendation (whether user will watch video)

Examples

- The Naive Bayes Algorithm
- Support Vector Machines
- etc.

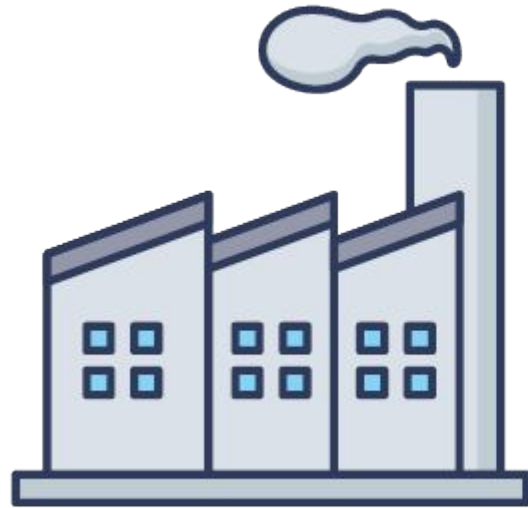
Algorithms

Unsupervised Learning

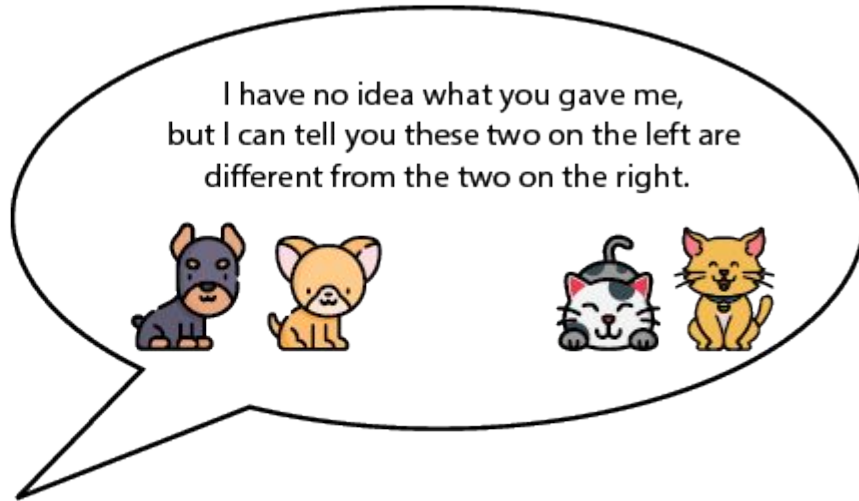
Extract as much as possible information from data.

Group the images based on similarity.

Separate the dog images from the cat images, or even group each of them by breed!



Unsupervised learning algorithm



Unsupervised Learning

- **Clustering algorithms** group data into clusters based on similarity;
- **Dimensionality reduction algorithms** simplify data and describe it with fewer features; **Main branches**
- **Generative algorithms** generate new data points that resemble the existing data.

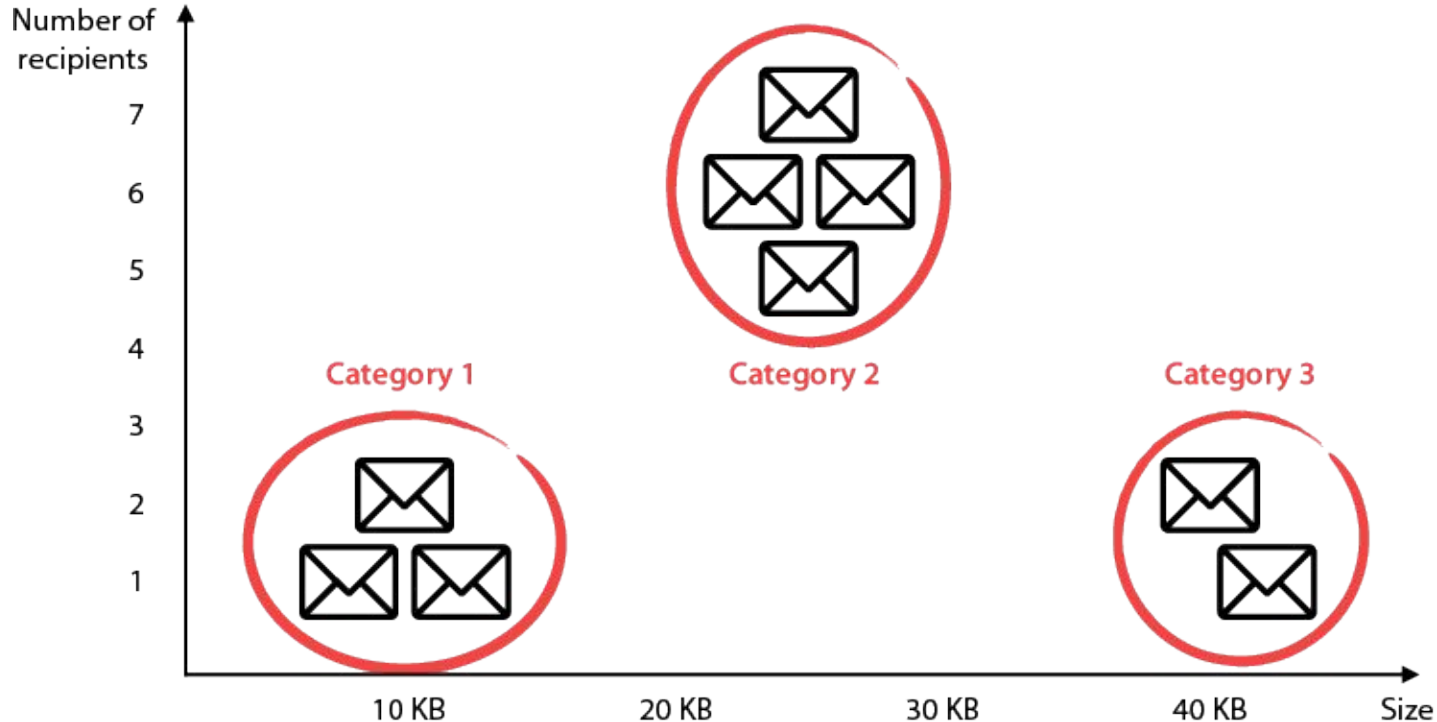
1. **Use unsupervised learning techniques on data to preprocess**
2. **Apply supervised learning methods more effectively**

- **K-means clustering**
- **Hierarchical clustering**
- **Density based spatial clustering**
- **Gaussian mixture models**
- **etc.**

Algorithms

Unsupervised Learning. Clustering algorithm

Example about spam emails. Group emails by the size, sender, links in email, attachments etc.



Unsupervised Learning. Dimensionality reduction.

Preprocessing step to simplify data prior to other techniques.

Selling houses dataset:

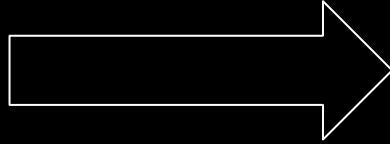
- Size
- Number of bedrooms
- Number of bathrooms
- Crime rate in the neighborhood
- Distance to the closest school

Unsupervised Learning. Dimensionality reduction.

Preprocessing step to simplify data prior to other techniques.

Selling houses dataset:

- Size
- Number of bedrooms
- Number of bathrooms
- ❑ Crime rate in the neighborhood
- ❑ Distance to the closest school

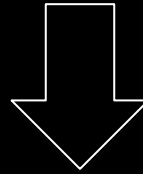


- General size
- ❑ Neighbourhood quality

Unsupervised Learning. Dimensionality reduction.

Preprocessing step to simplify data prior to other techniques. Selling houses dataset:

Size	Number of bedrooms	Number of bathrooms	Crime rate in the neighborhood	Distance to the closest school



General size	Neighbourhood quality

Unsupervised Learning. Dimensionality reduction.

Self study:

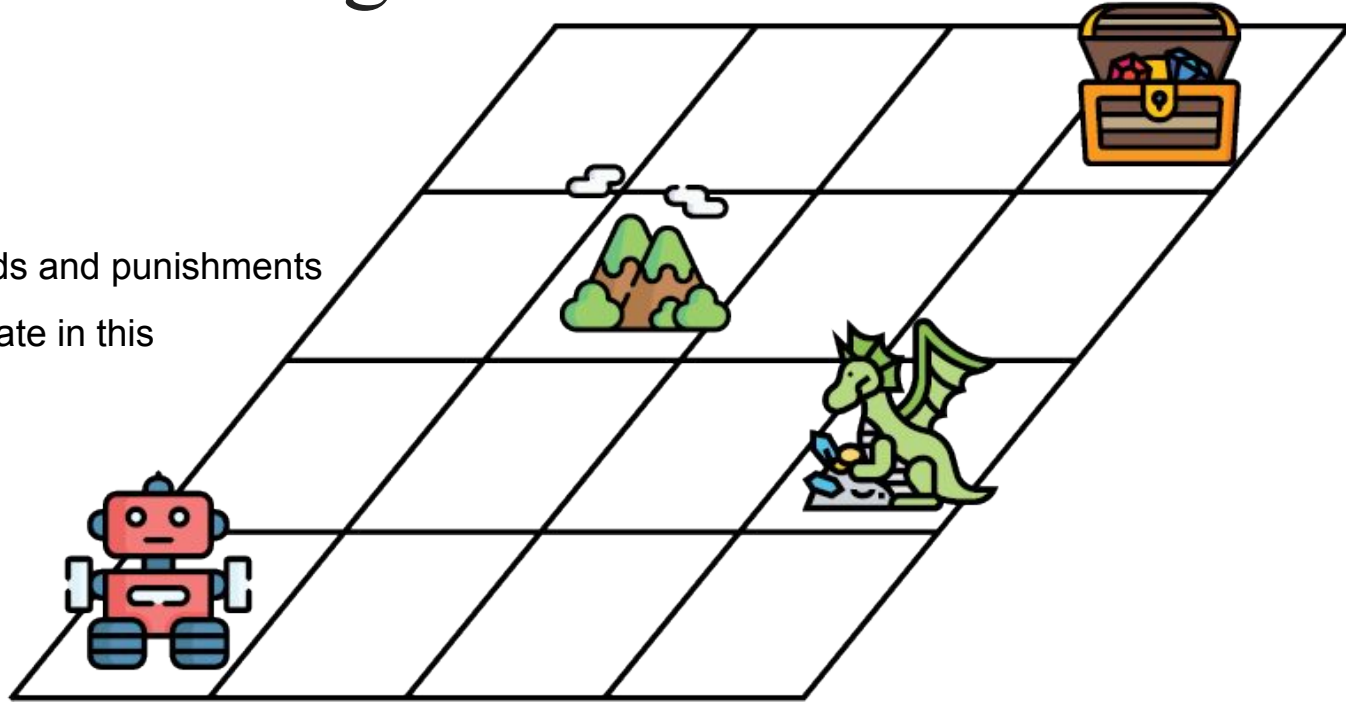
- Matrix factorization
- Singular value decomposition

Reinforcement Learning

INPUT:

- No data.
- An environment with rewards and punishments
- An agent with goal to navigate in this environment.

OUTPUT: Perform a task.



Games: teach computers how to win at Go, Chess, Breakout, Super Mario etc. games.

Robotics: help robots picking up boxes, cleaning a room, or even dancing!

Self-driving cars: path planning or behaving in particular environments.

Machine Learning

- Supervised learning
 - Regression
 - Classification
- Unsupervised learning
 - Clustering
 - Dimensionality reduction
 - Generative ML
- Reinforcement learning

- Predicting house prices based on the house's size, number of rooms, and location
- Predicting today's stock market prices based on yesterday's prices and other factors of the market
- Detecting spam and non-spam emails based on the words in the e-mail and the sender
- Recognizing images as faces or animals, based on the pixels in the image
- Processing long text documents and outputting a summary
- Recommending videos or movies to a user (e.g., on YouTube or Netflix)
- Building chatbots that interact with humans and answer questions
- Training self-driving cars to navigate a city by themselves
- Diagnosing patients as sick or healthy
- Predicting whether a credit card transaction is legitimate or fraudulent
- Segmenting the market into similar groups based on location, acquisitive power, and interests
- Playing games like Chess or Go

Summary

- Major types of machine learning: supervised learning, unsupervised learning, and reinforcement learning.
- Data can be labeled or unlabeled. Labeled data contains a special feature that we aim to predict.
- Supervised learning is used on labeled data and consists of building models that predict the labels for unseen data.
- Unsupervised learning is used on unlabeled data and consists of algorithms that simplify our data without losing a lot of information. Unsupervised learning is often used as a preprocessing step.
- Two common types of supervised learning algorithms are called regression and classification.
- Regression models are those in which the answer is any number.
- Classification models are those in which the answer is of a type or a class.
- Two common types of unsupervised learning algorithms are clustering and dimensionality reduction.
- Clustering is used to group data into similar clusters to extract information or make it easier to handle.
- Dimensionality reduction is a way to simplify our data, by joining certain similar features and losing as little information as possible.
- Matrix factorization and singular value decomposition are other algorithms that can simplify our data by reducing both the number of rows and columns.
- Generative machine learning is an innovative type of unsupervised learning, consisting of generating data that is similar to our dataset. Generative models can paint realistic faces, compose music, and write poetry.
- Reinforcement learning is a type of machine learning in which an agent must navigate an environment and reach a goal. It is extensively used in many cutting-edge applications.