Day_01

Video 01: Introduction to Machine Learning (ML)

Why ML?

+ Used in 5G, autonomous driving, sports analytics, e-commerce, and more.
+ Forbes listed AI/ML as a top digital transformation trend. Google Trends shows a boom in ML/DL/AI searches from 2016 onwards.

Where is ML applied?

+ Vision, speech, text, understanding the brain, stock markets, biological/chemistry experiments, sports analytics, e-commerce recommendations.

What is NOT ML?

+ Not a procedural/algorithmic approach (tax calculation)
+ Not memorization: memorizing 2000 tree pictures doesn't mean you understand what a tree is; you won't generalize to new pictures
+ Not magic: it's math

What is ML?

+ Data is the secret sauce (not just algorithms)
+ Generalization: doing well on unseen data, not just data you've trained on
+ Math: mathematical models that learn from data

My definition: *Machine learning is the design of mathematical models that learns from data to perform a defined task and generalize those patterns to new data drawn from the same distribution.*

Examples of ML problems

+ Spam vs non-spam, forecasting rainfall, movie recommendations, friend suggestions (LinkedIn), voice/instrument separation, grouping faces on phone, robot navigation, and stock market prediction.

Video 02: Paradigms of ML

Paradigms

Supervised Learning: given X (features) + Y (labels/correct answer), has a goal. Learn to map X → Y.

+ Classification: labels are finite
  - Binary (spam vs non-spam)
  - Multi-class (digit recognition, 0–9)

  - Ordinal / Ranking (movie recommendations in order)
  - Structured learning (predicting trees/structured outputs)
+ Regression: predicting a continuous output (rainfall amount)
+ Ranking / Link prediction (friend suggestions)

Unsupervised Learning: only X given, no labels. Self-exploratory, learn/generalise structure/patterns.

+ Clustering / Grouping (grouping phone pictures by face)
+ Representation learning (learning better ways to represent data)
+ Voice/instrument separation

Sequential Learning: online, learn over time through feedback (not covered in this course)

+ Online learning: full information feedback (stock market prediction)
+ Multi-armed bandits: partial feedback (treatment recommendation)
+ Reinforcement learning: learn a policy via environment feedback (robot navigation)

Examples mapped to paradigms

+ Spam vs non-spam (binary classification)
+ Forecasting rainfall (regression)
+ Movie recommendations (ordinal/ranking)
+ Friend suggestions (ink prediction)
+ Voice/instrument separation (unsupervised)
+ Grouping pictures (clustering)
+ Robot navigation (reinforcement learning)
+ Stock market (online learning)

Prerequisites

+ Structure: linear algebra
+ Uncertainty: probability
+ Data to decisions: optimization
+ Also: statistics and basic calculus.

Course Roadmap

Unsupervised learning (representation learning, clustering, estimation) → supervised learning (basic to advanced algorithms) → advanced supervised learning topics.

References

+ Linear Algebra by Gilbert Strang
+ Probability by Sheldon Ross (A First Course)
+ Main reference by Pattern Recognition and ML by Christopher Bishop (PRML)
+ Also: Mathematics for Machine Learning (free soft copy available)

Day_01

<u>Video 03: Representation Learning (Unsupervised Learning)</u>

UNSUPERVISED LEARNING

↳ Representation learning

Goal: Given a set of "data points", "understand" something "useful" about them

Datapoint → vectors in $\mathbb{R}^d$ $\begin{bmatrix} \text{height} \\ \text{weight} \\ \text{age} \end{bmatrix} \in \mathbb{R}^3$

Running theme → "Comprehension is compression" — George Chaitin

↳ understanding
→ learning

Problem

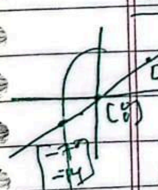
Input: $\{x_1, x_2, \ldots x_n\}$ $x_i \in \mathbb{R}^d \leftarrow$ # Features

Output: Some "compressed" representation of the dataset

Example

$$\left\{ \overset{x_1}{\begin{bmatrix} -7 \\ -14 \end{bmatrix}}, \overset{x_2}{\begin{bmatrix} 2.5 \\ 5 \end{bmatrix}}, \overset{x_3}{\begin{bmatrix} 0.5 \\ 1 \end{bmatrix}}, \overset{x_4}{\begin{bmatrix} 0 \\ 0 \end{bmatrix}} \right\}$$

Question How many real numbers are needed to store this dataset .. $\boxed{8}$ ans



Representative $\begin{bmatrix} 1 \\ 2 \end{bmatrix} \in \mathbb{R}^2$ | co-efficients $\{-7, 2.5, 0.5, 0\}$, $\boxed{6}$ → ans

Note

→ Using representative & co-efficients can "RECONSTRUCT" the dataset exactly.

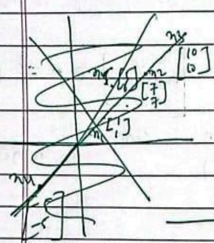
→ Any vector along the line in the graph can be chosen as a representative [except $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$]

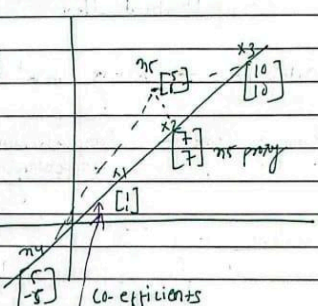
Day_01

orignal:

\# real numbers $= dn$

\# real numbers in compressed representation $= d + n$





Representatile

co-efficients

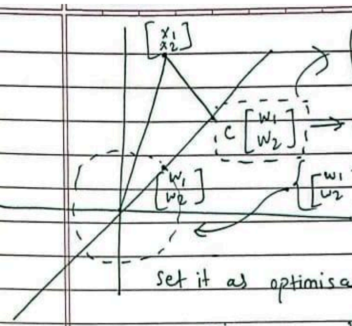
$\left\{ \begin{bmatrix}1\\0\end{bmatrix} \begin{bmatrix}0\\1\end{bmatrix} \right\}$

$\{ (-5,5), (1,), (7,7), (5,1), (10,0] $

$x_3 = 10 \begin{bmatrix}1\\0\end{bmatrix} \begin{bmatrix}0\\1\end{bmatrix} = \begin{bmatrix}10\\10\end{bmatrix}$

Q: Who can "pretend" to be a proxy for x_5 along the blue line?

Ans: Projection of x_5 on / line



$\left(\dfrac{x_1 w_1 + x_2 w_2}{w_1^2 + w_2^2}\right) \begin{bmatrix}w_1\\w_2\end{bmatrix} = \left(\dfrac{x^T w}{\|w\|^2}\right) \begin{bmatrix}w_1\\w_2\end{bmatrix}$

$c \begin{bmatrix}w_1\\w_2\end{bmatrix} \rightarrow$ projection of $\begin{bmatrix}x_1\\x_2\end{bmatrix}$

$\left\{ \begin{bmatrix}w_1\\w_2\end{bmatrix} : w_1^2 + w_2^2 = 1 \right\}$

Set it as optimisation problem.

$\min_c \ \text{length}^2 (\text{error vector})$

↳ it What should be added $c \begin{bmatrix}w_1\\w_2\end{bmatrix}$ to get $\begin{bmatrix}x_1\\x_2\end{bmatrix}$

C depends on $\begin{bmatrix}w_1\\w_2\end{bmatrix}$ & $\begin{bmatrix}x_1\\x_2\end{bmatrix}$

$= \begin{bmatrix}x_1 - cw_1\\x_2 \ \ cw_2\end{bmatrix}$

$\min_c \ (x_1 - cw_1)^2 + (x_2 - cw_2)^2$

$c^* = \left(\dfrac{x_2 w_1 + x_2 w_2}{w_1^2 + w_2^2}\right)$ (scalar)

Inner product/ dot product of x and w

$\text{length}^2 \left(\begin{bmatrix}w_1\\w_2\end{bmatrix} \right)$

Note : can pick pick $\begin{bmatrix}w_1\\w_2\end{bmatrix}$ such that length $\left(\begin{bmatrix}w_1\\w_2\end{bmatrix} \right) = 1$

$\Rightarrow c^* = (x^T w) \begin{bmatrix}w_1\\w_2\end{bmatrix}$

End of Note
27 Feb, 2026