Georgia Tech

Methods & Applications of Machine Learning

Mohsen Moghaddam, Ph.D.

Gary C. Butler Family Associate Professor
H. Milton Stewart School of Industrial and Systems Engineering
George W. Woodruff School of Mechanical Engineering
Georgia Institute of Technology

SAIL
Symbiotic & Augmented Intelligence Laboratory

Spring 2026

Course Description and Objectives

Description

Introduction to machine learning (ML) with emphasis on foundations and ISyE applications. Topics include clustering, dimensionality reduction, classification (such as Bayes classifier, logistic regression), optimization for ML, neural networks, feature selection, anomaly detection, ensemble methods such as random forest and boosting, and model assessment.

Objective

This course develops an understanding of core ML methods and their practical use in ISyE contexts. Upon completion, students will be able to: **Gain a thorough understanding of ML algorithms, use ML algorithms and choose the correct ones, and gain experience with analyzing real data.**

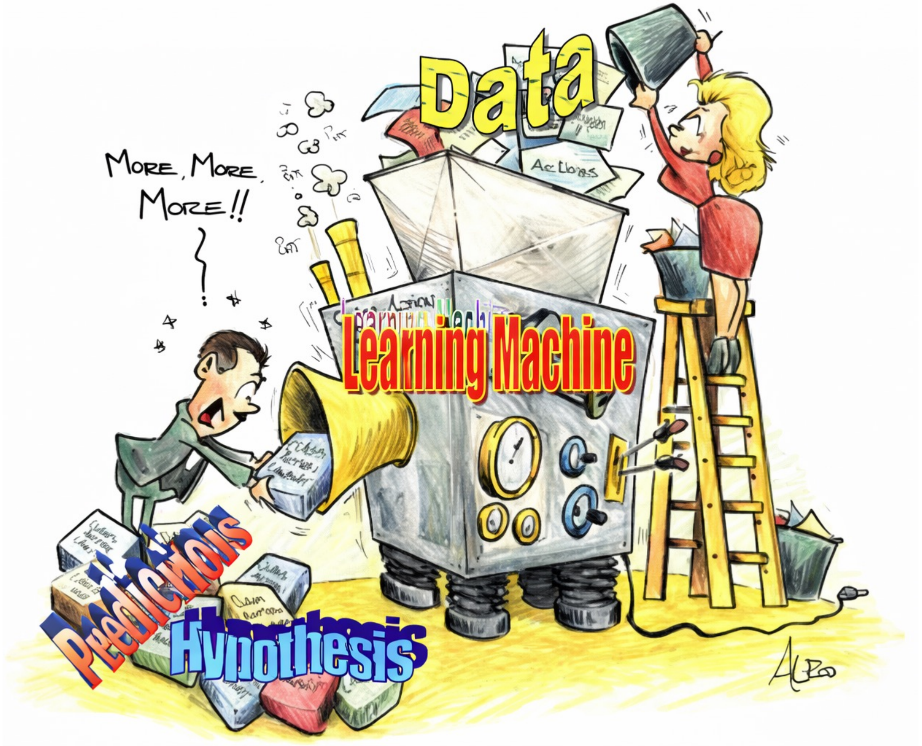
Why Machine Learning Matters

What is Machine Learning?

A set of methods that can **automatically detect patterns** in data and use them to predict (or generate) future data

Machine learning provides scalable, automated decision rules when

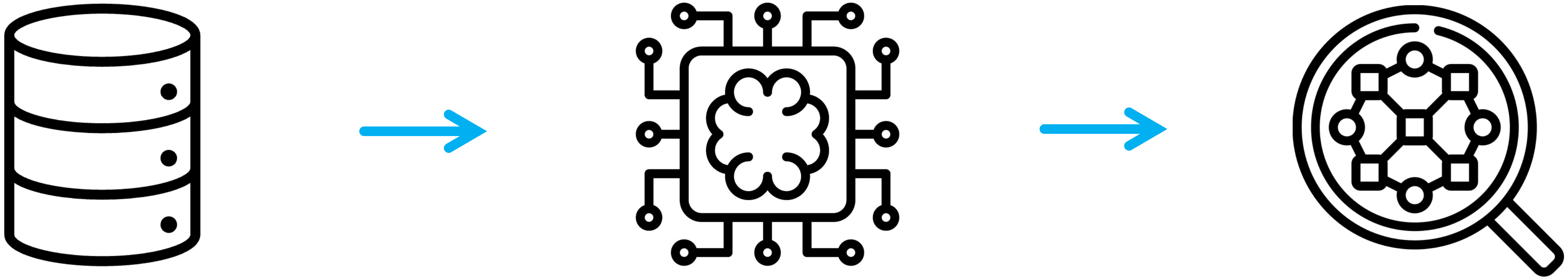
- System dynamics are complex or unknown
- Data is abundant but first-principles models are incomplete
- Decisions must adapt over time
- Engineered systems must learn from data



ML gives you tools to optimize when models are unknown or data is massive

A Common ML Problem Template

- **Inputs:** data collected from sensors, logs, users, networks
- **Learning paradigm:** supervised / unsupervised / reinforcement learning
- **Outputs:** decisions, groupings, rankings, predictions (regression/classification)



Inputs

High-dimensional data (images, pixels, text, numbers)

Learning Paradigm

Unsupervised learning / representation learning (clustering, embeddings, dimensionality reduction)

Outputs

Structured discovery (groups, communities, low-dimensional visualizations)

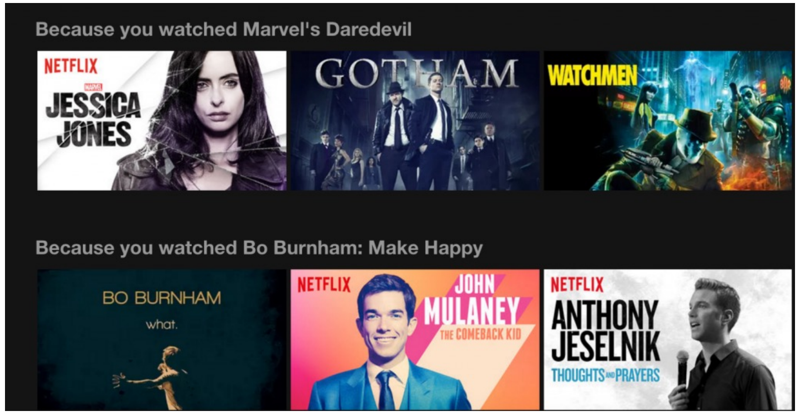
Growing Relevance to Science & Engineering

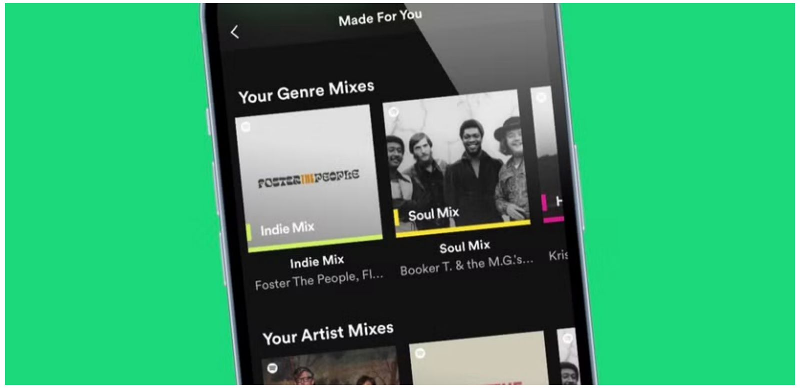
Impact across industries:

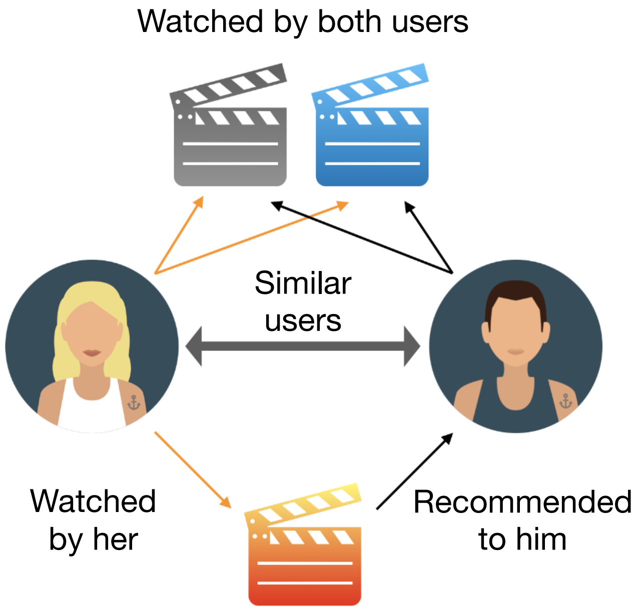
- **Manufacturing:** defect detection, predictive maintenance
- **Supply chains:** demand forecasting, inventory optimization
- **Healthcare operations:** patient flow, risk stratification
- **Energy systems:** load forecasting, anomaly detection
- **Transportation:** routing, traffic prediction, autonomous systems

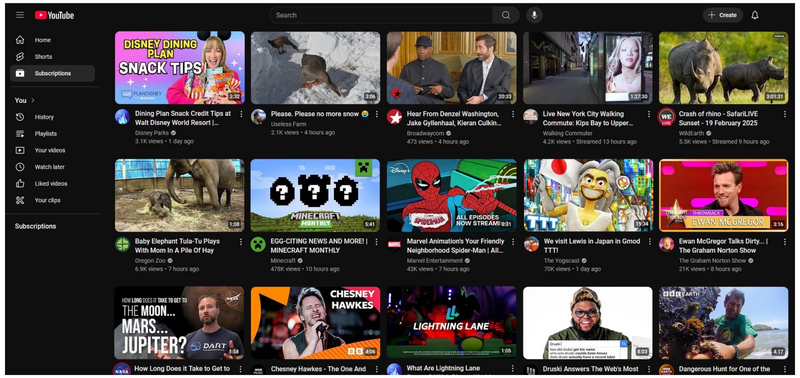
ML is now foundational engineering tools used across nearly every industry

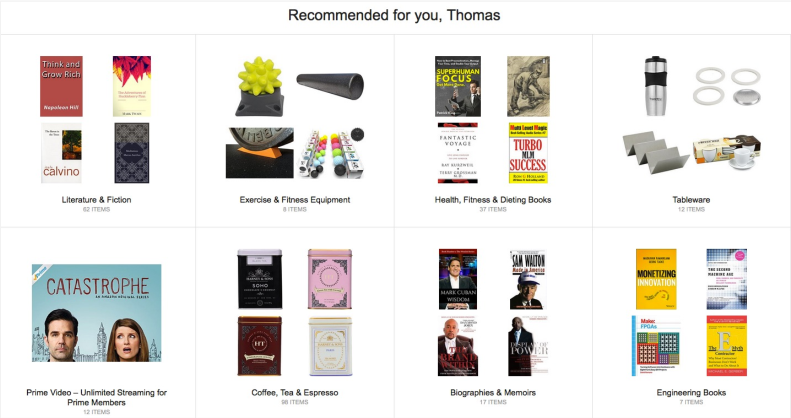
Generate Personalized Recommendations





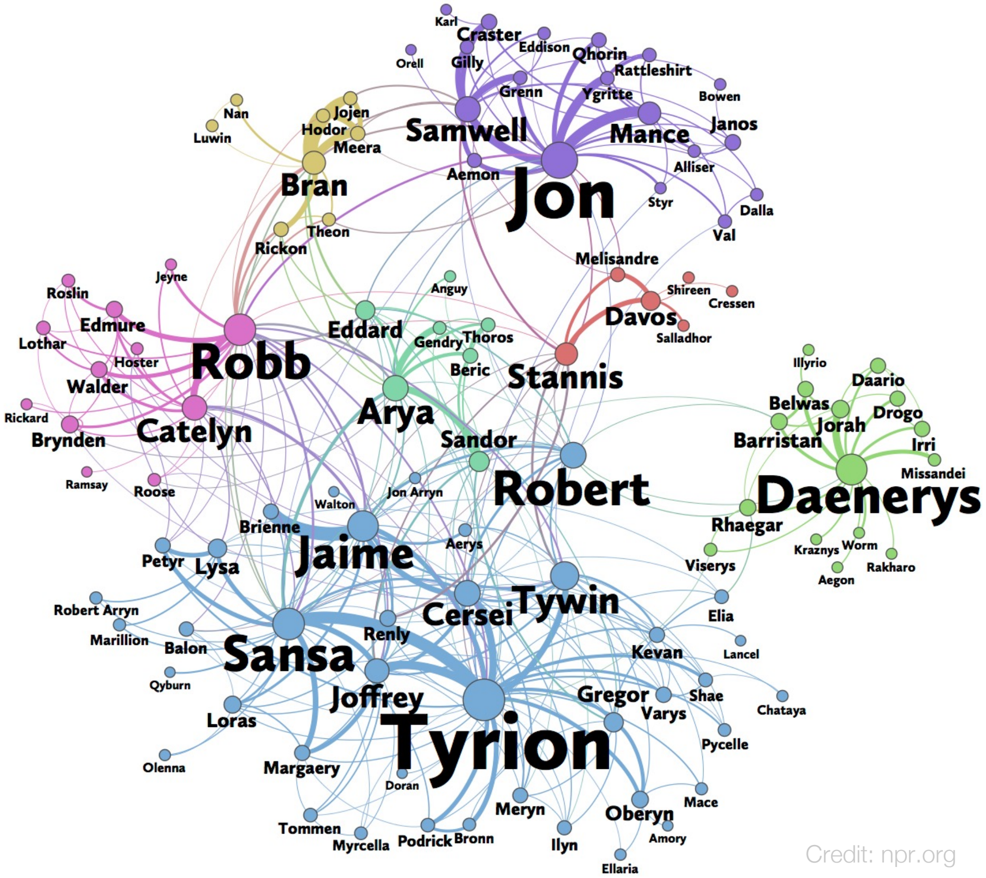






ML enables personalized recommendations by predicting user preferences

Find Community in Social Networks



Credit: npr.org

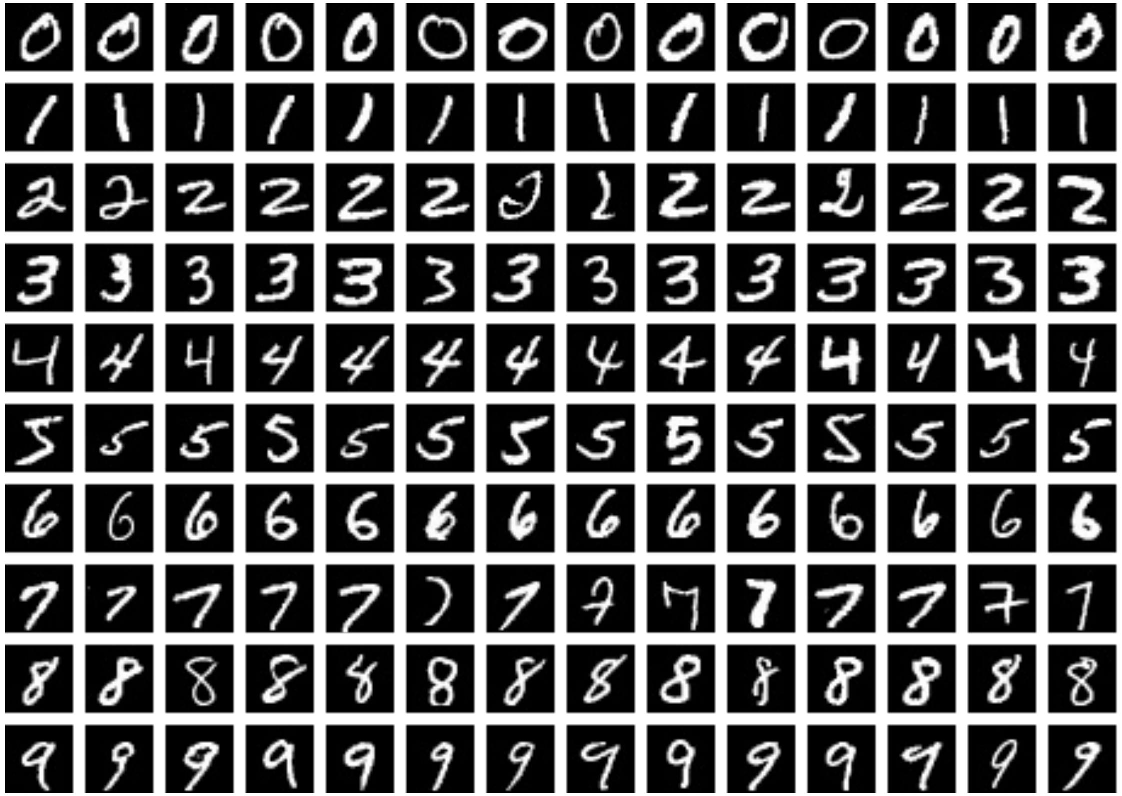
ML identifies hidden groups and relationships in networked data

Organize Images

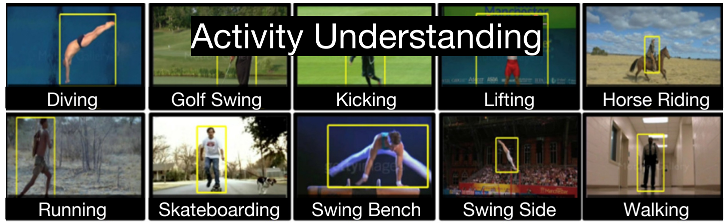
Partition data into groups such that objects within a group are more similar to each other than to objects in other groups

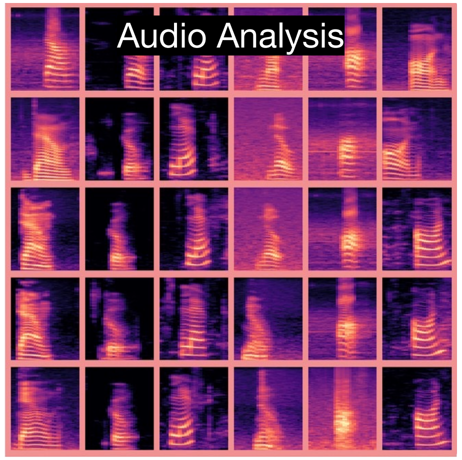
ML discovers structure in large, unstructured image datasets

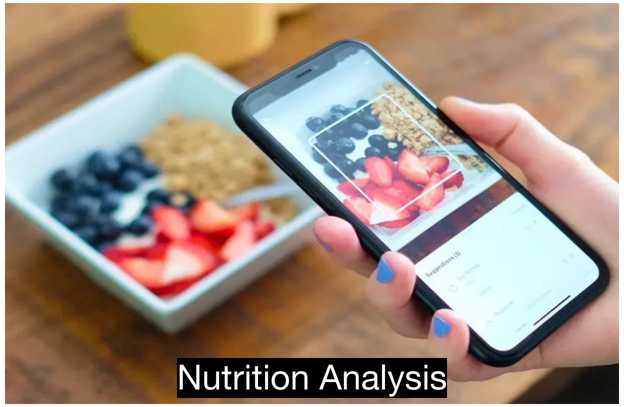
Classify Images

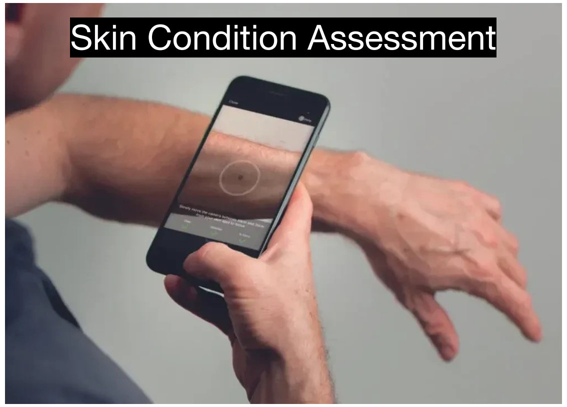


MNIST Dataset (Handwritten Digit Classification)



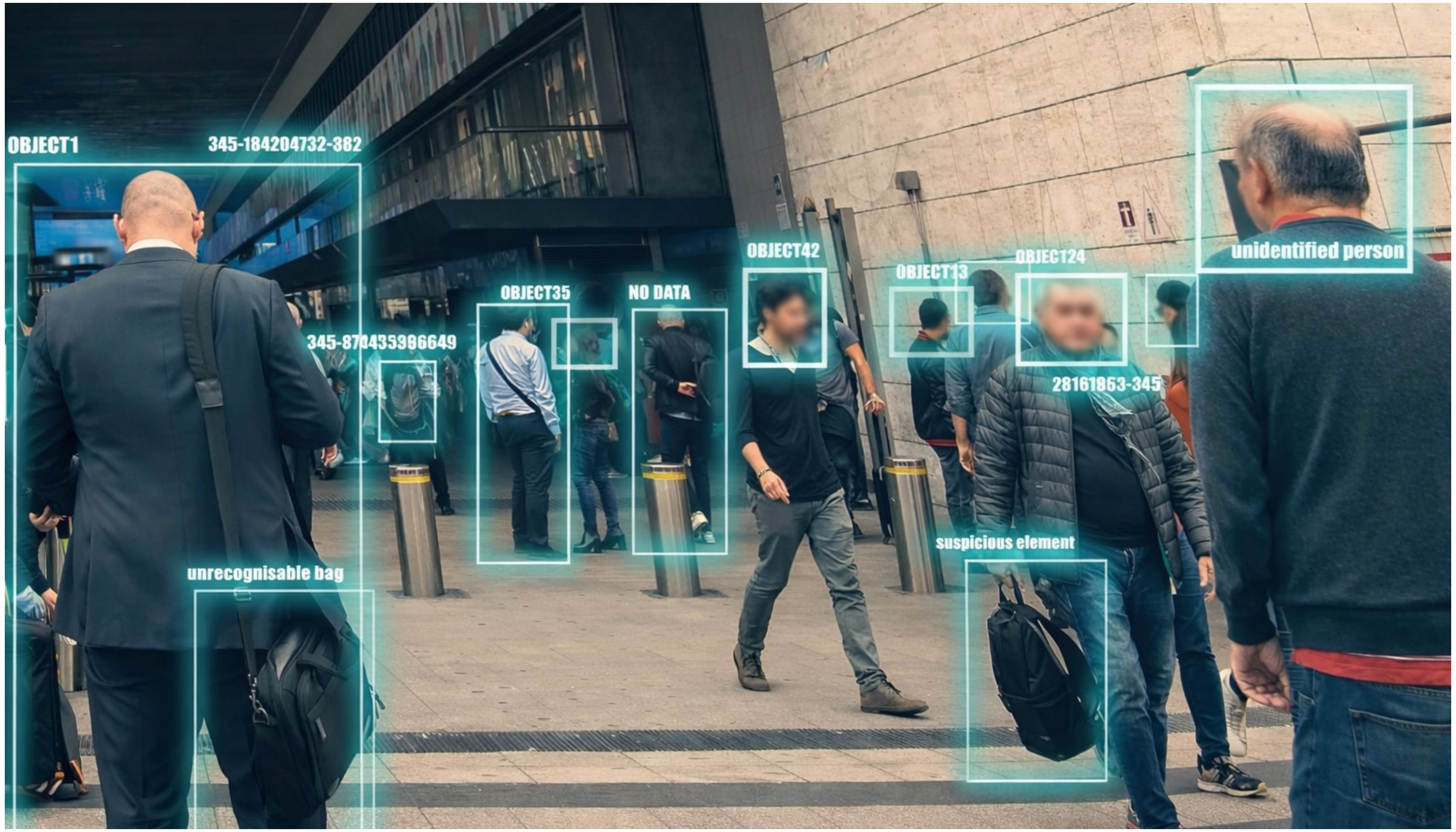






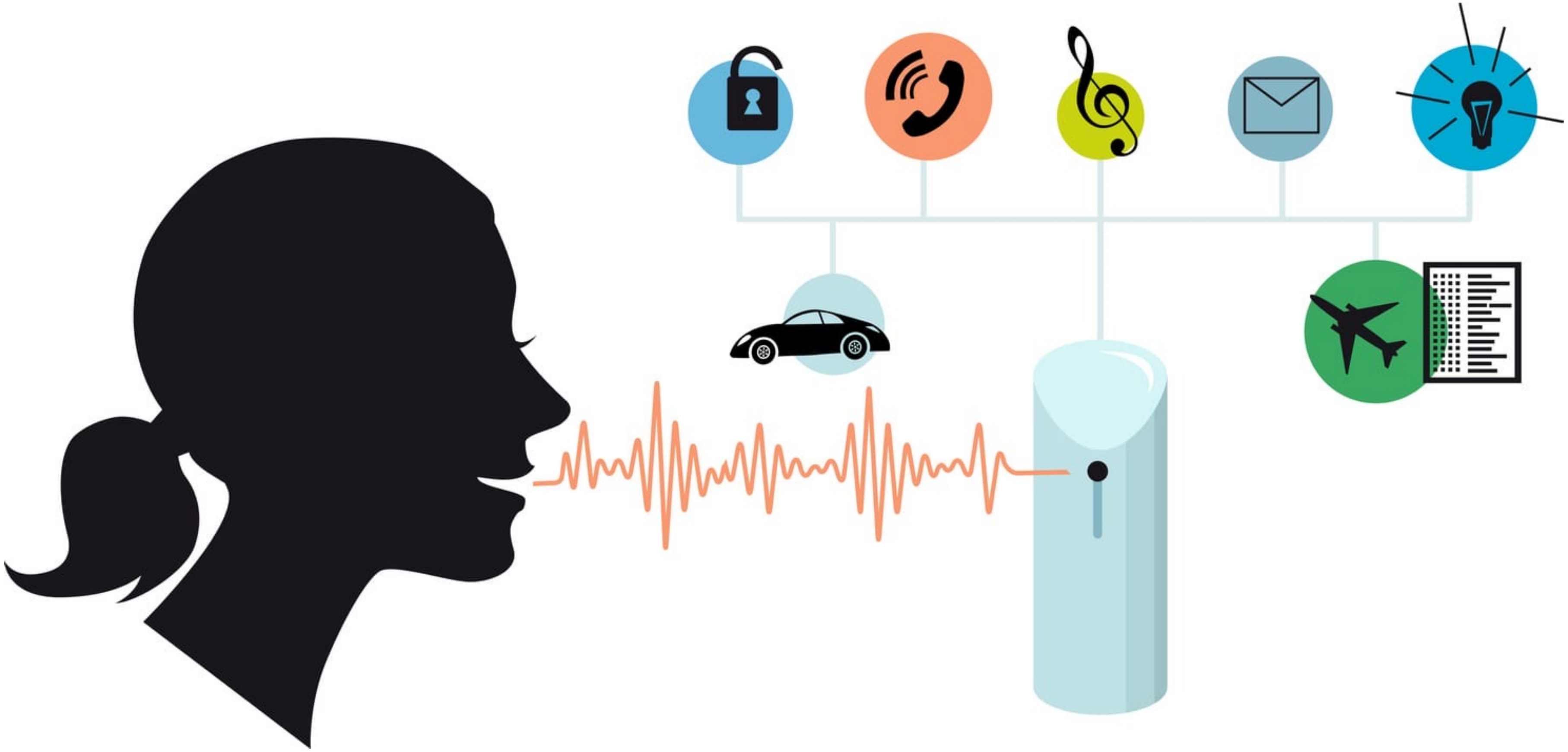
ML automatically assigns labels to data based on learned patterns

Recognize and Localize Objects in Images



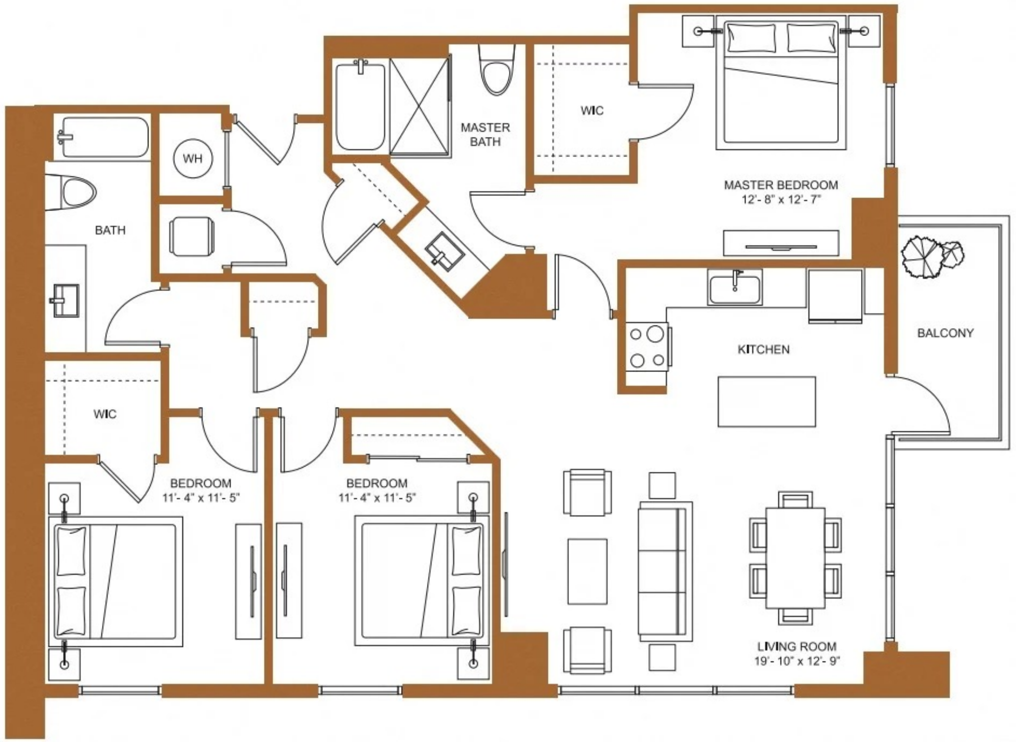
ML recognizes and locates specific objects of interest within raw data

Recognize Speech



ML converts complex sensor signals into structured, actionable information

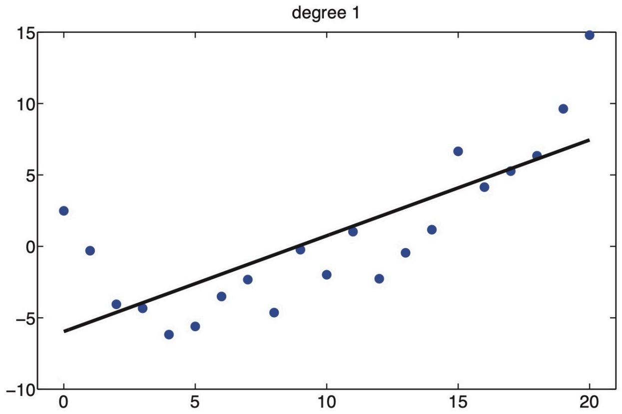
Identify Key Predictors

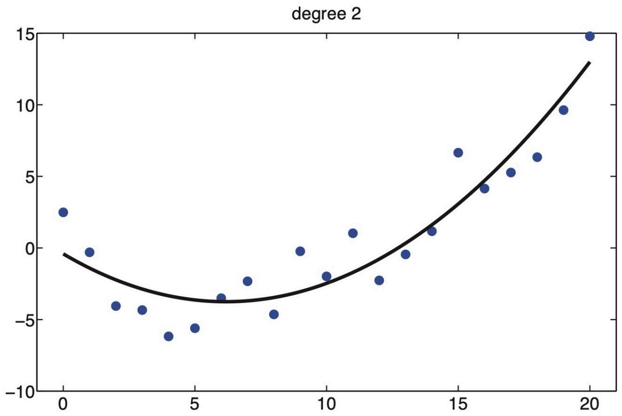


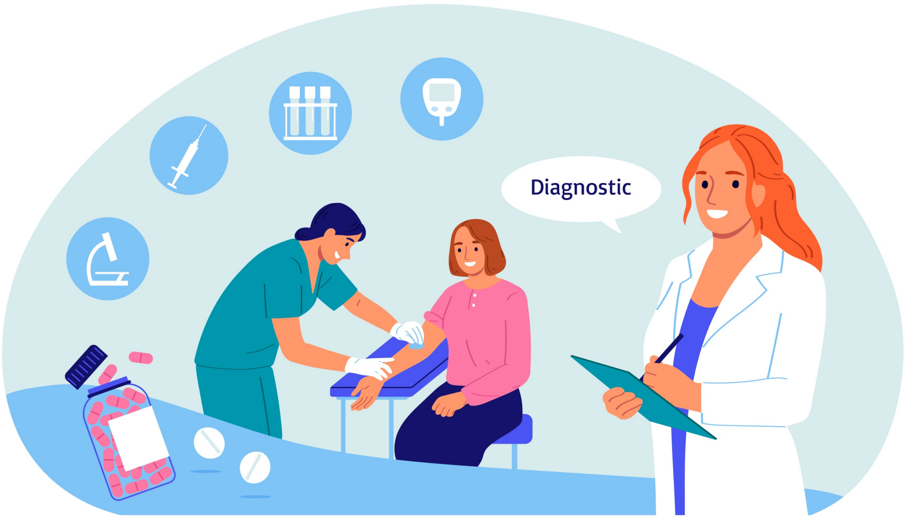
Square ft.? # Bedrooms? # Bathrooms? Distance to Work?

ML identifies key variables that help accurately predict housing prices

Advance Medical Diagnosis





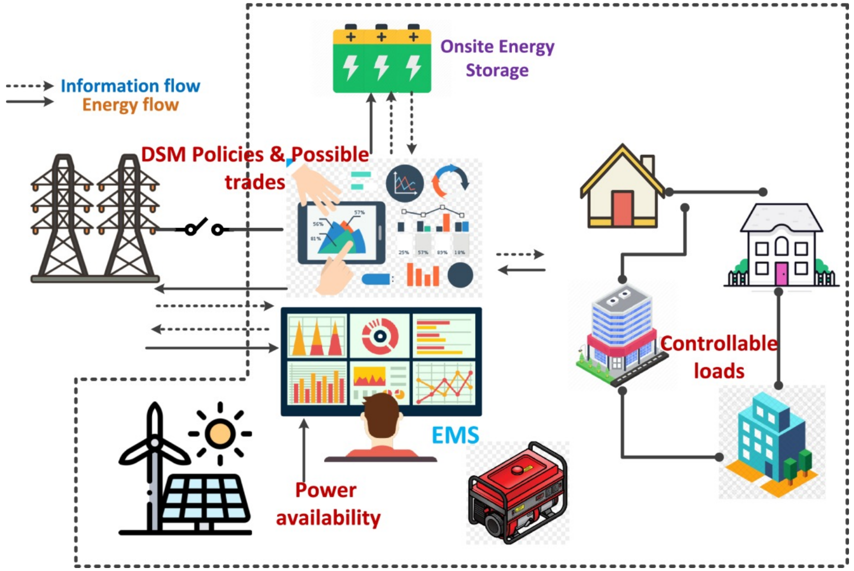


Diabetes Data

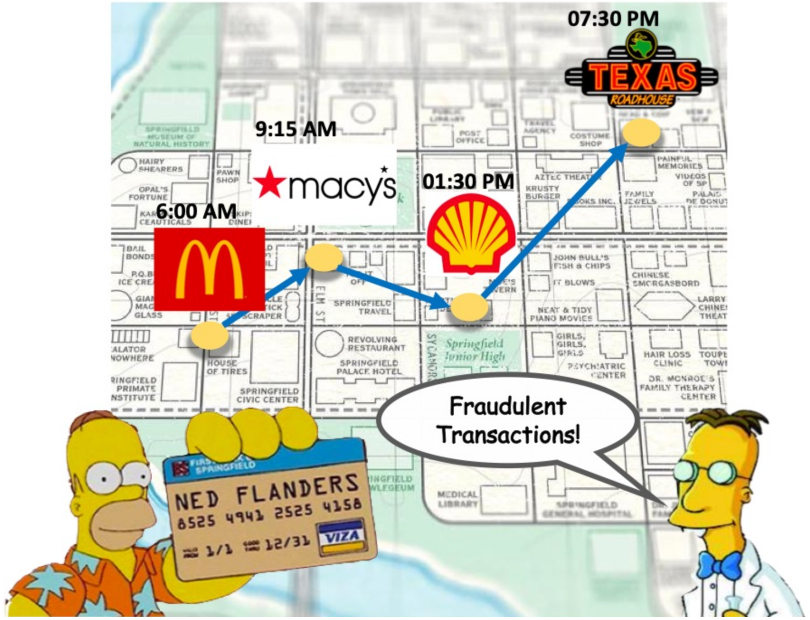
- # Diabetes Patients
- Baseline predictors
- Age, Sex, Body Mass Index
- Average Blood Pressure
- Six Blood Serum Measurements

ML predicts measure of disease progression such as diabetes

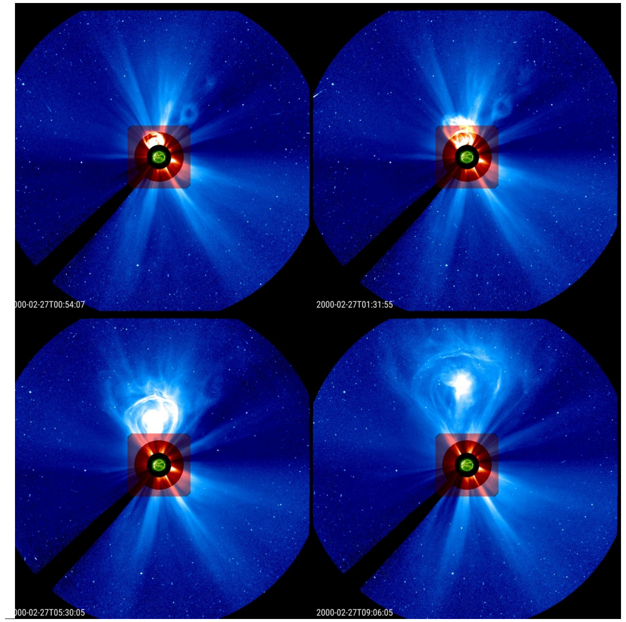
Detect Anomalies



Power Network Monitoring



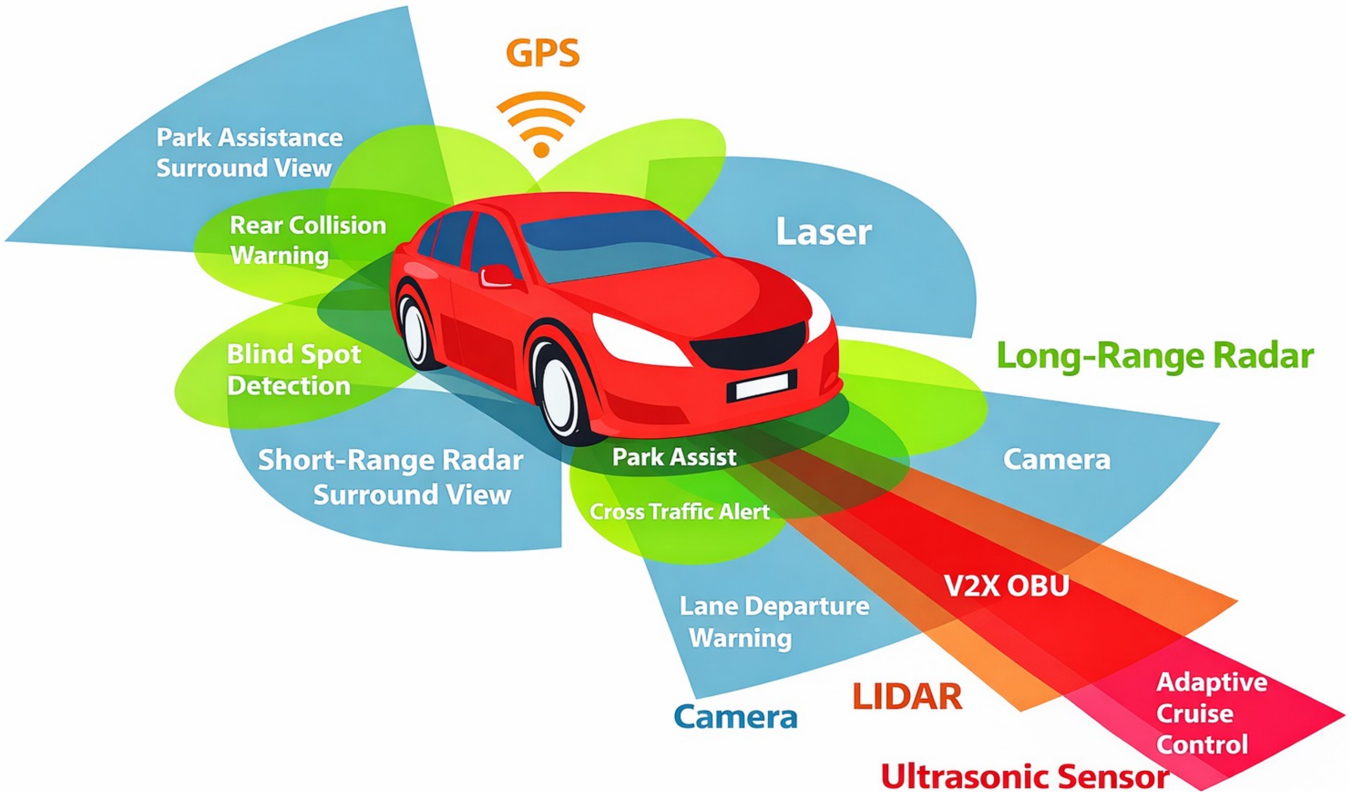
Credit Card Fraud Detection



Solar Flare Detection

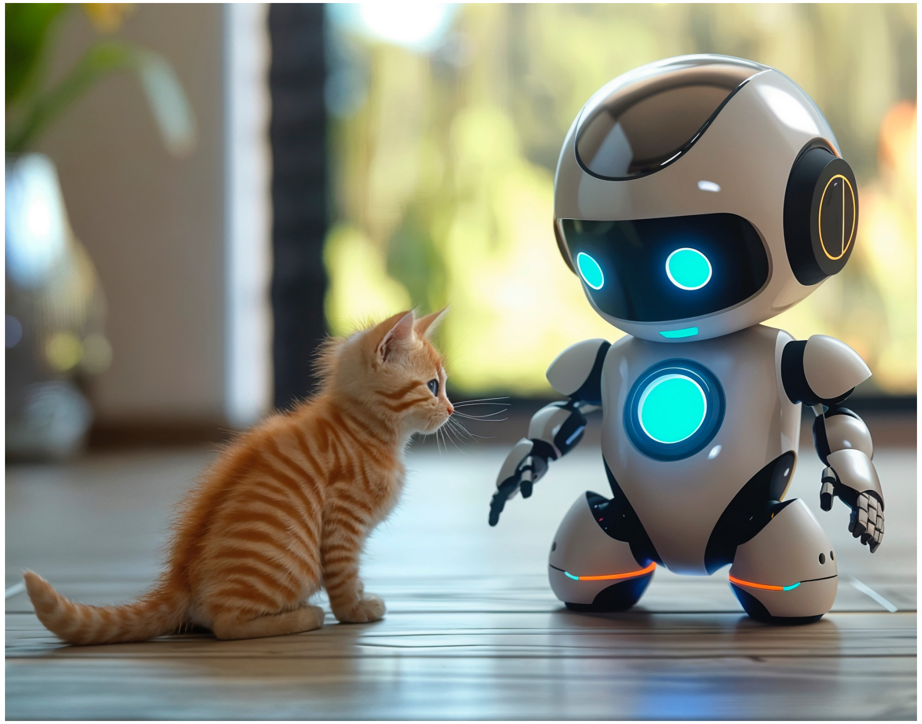
ML detects rare, abnormal, or anomalous behavior in complex systems

Enable Safe Autonomous Robots



ML detects abnormal scenarios for safe navigation of autonomous robots

Generate Text, Images, and Videos


“The sun slipped behind the hills, painting the sky in soft oranges and purples, while a curious little robot quietly wondered what it would dream about tonight.” 🌅🤖

ML generates realistic content by learning the underlying data distribution

What You'll Learn in ISYE 4600

Unsupervised Learning

Explore unsupervised learning techniques for discovering hidden patterns and structures in data without labeled examples

Clustering (W2)
Principal Component Analysis (W3)
Recommender Systems (W4)

Model Selection, Evaluation, & Optimization

Build a solid foundation in machine learning concepts, theory, and mathematical principles that underpin all ML algorithms

Feature Selection (W5)
Bias–Variance Tradeoff & Cross Validation (W6)
Basic Optimization (W7)

Supervised Learning

Master supervised learning algorithms for classification and regression tasks using labeled training data

Classification (W8)
Support Vector Machine (W9)
Neural Networks (W10)

Anomaly Detection & Ensemble Learning

Learn advanced ensemble techniques that combine multiple models to achieve superior predictive performance

Anomaly Detection (W11)
Boosting Algorithms (W12)
Decision Trees & Random Forests (W13)

Prerequisites

Prerequisites for Machine Learning

Probability Theory

- Random variables and probability distributions
- Probability densities, marginalization, and conditioning

Statistics

- Descriptive statistics: mean and variance
- Parameter estimation, including maximum likelihood estimation

Linear Algebra

- Vectors and matrices
- Matrix operations (multiplication, inversion)
- Eigenvalues and eigendecomposition

Optimization

- Fundamentals of convex optimization
- Core concepts will be introduced during the course

Programming

- Programming experience (e.g., Python or equivalent)
- Basic data structures
- Computational complexity and efficiency

Example: ML for Apartment Hunting

Suppose you are to move to a new city and want to find the most reasonably priced apartment satisfying your needs

Rent ($)	Living area (ft^2)	Location	# bath	# bedroom
600	230	Midtown	1	1
1000	506	Buckhead	2	2
1100	433	Midtown	1	2
500	109	Downtown	1	1
⋮	⋮	⋮	⋮	⋮
?	**150**	**Midtown**	**2**	**1**
?	**270**	**Downtown**	**1**	**1.5**

What skills do you need to solve this problem using machine learning?

Linear Regression Model

We model the output as a **linear function of the input features plus noise**—this simple model forms the foundation of many machine learning methods

$$y = \theta_0 + \theta_1 x_1 + \cdots + \theta_n x_n + \epsilon$$

Probability

where the **residual error** ϵ follows a Gaussian distribution $\epsilon \sim \mathcal{N}(0, \sigma^2)$

Let $\theta = (\theta_0, \theta_1, \dots, \theta_n)^\top$ and augment data by one dimension

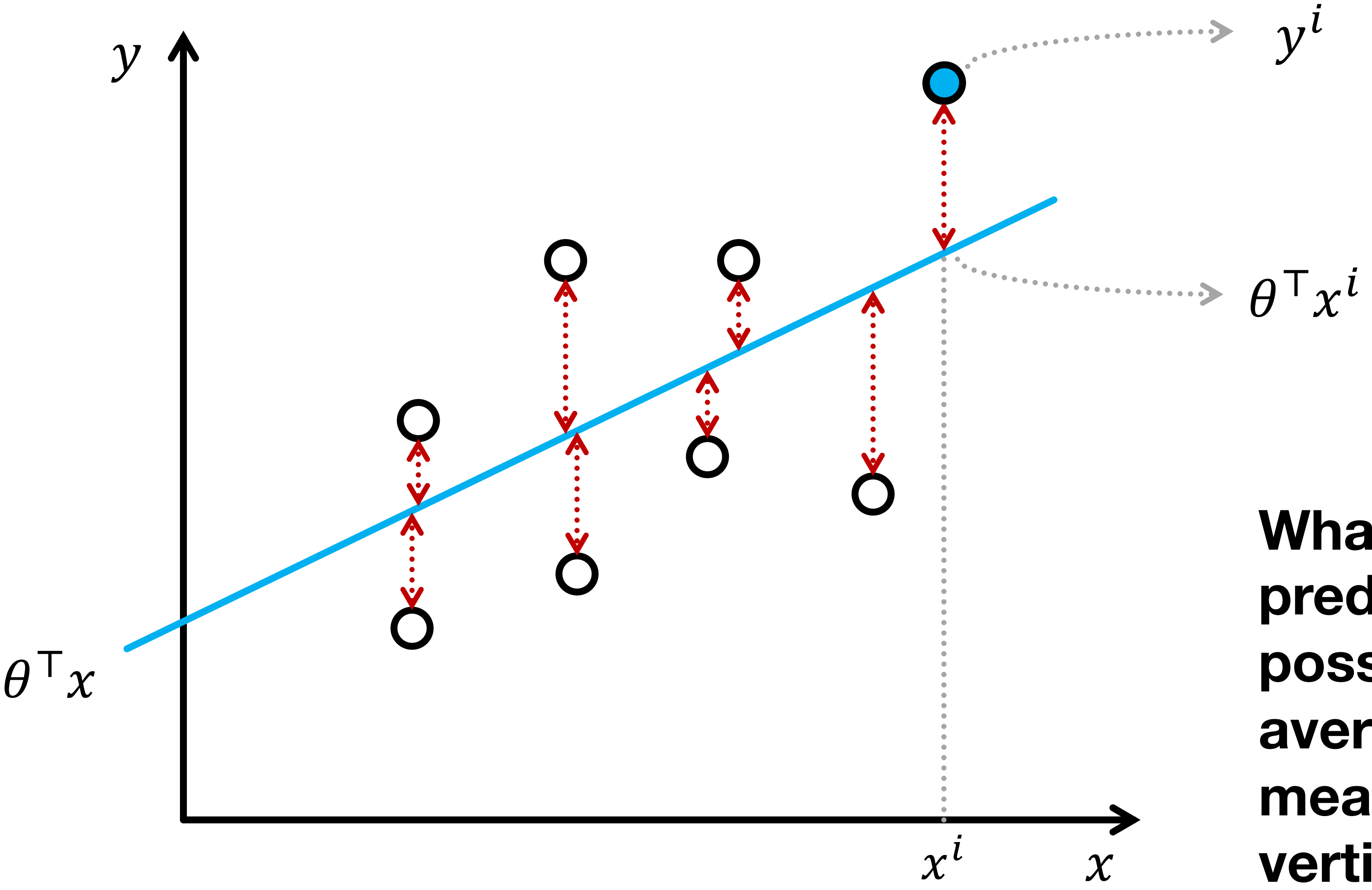
Linear Algebra

$$x \leftarrow (1, x)^\top$$

m data points
n features
$y \in \mathbb{R}^m$
$\theta \in \mathbb{R}^{n+1}$
$x \in \mathbb{R}^{(n+1)\times m}$

Then $y = \theta^\top x + \epsilon$

Linear Regression Model (cont.)



What “line” makes our predictions as close as possible to the data, on average, when closeness is measured by squared vertical error?

Mean Squared Error (MSE) Estimation

Given m training data, we estimate the parameters by minimizing the average squared prediction error—this is an **optimization problem**:

$$\hat{\theta} = \arg\min_{\theta} L(\theta) \qquad\qquad L(\theta) = \frac{1}{m}\sum_{i=1}^{m}\left(y^i - \theta^\top x^i\right)^2$$

Optimization

Linear Algebra

Set the **gradient** to zero to find the optimal solution:

$$\frac{\partial L(\theta)}{\partial \theta} = -\frac{2}{m}\sum_{i=1}^{m}\left(y^i - \theta^\top x^i\right)x^i = -\frac{2}{m}\sum_{i=1}^{m} y^i x^i + \frac{2}{m}\sum_{i=1}^{m} x^i x^{i\top}\theta = 0$$

Optimization

Linear Algebra

Statistics

Matrix Form and Gradient Descent

Using matrix notation simplifies computation and reveals computational challenges

$$X = (x^1, x^2, \dots, x^m) \qquad\qquad Y = (y^1, y^2, \dots, y^m)^\top$$

Gradient:

Linear Algebra

$$\frac{\partial L(\theta)}{\partial \theta} = -\frac{2}{m} X y + \frac{2}{m} X X^\top \theta \qquad \Longrightarrow \qquad \hat{\theta} = (XX^\top)^{-1} X y$$

Matrix inversion $\hat{\theta} = (XX^\top)^{-1} X y$ can be **expensive for large datasets**, so we use iterative optimization—**gradient descent** update:

Programming

$$\hat{\theta}^{t+1} = \hat{\theta}^{t} + \frac{\alpha}{m} \sum_{i=1}^{m} \left(y^i - \hat{\theta}^{t\top} x^i\right) x^i$$

Probabilistic Interpretation of MSE

Linear regression can be interpreted as a **probabilistic model**—assuming the noise term ϵ in linear function $y = \theta^\top x + \epsilon$ follows a Gaussian $\epsilon \sim \mathcal{N}(0, \sigma^2)$

$$p(y^i \mid x^i; \theta) = p(\epsilon = y^i - \theta^\top x^i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y^i - \theta^\top x^i)^2}{2\sigma^2}\right)$$

→ *Instead of saying "$y = \theta^\top x$," we say "y is a random variable whose distribution is a bell curve centered at $\theta^\top x$"—that is, we predict a distribution not a single number*

Assuming independence across samples, the **likelihood** is:

$$\boxed{\text{Probability}} \rightarrow \quad L(\theta) = \prod_{i=1}^{m} p(y^i \mid x^i; \theta)$$

Probabilistic Interpretation of MSE (cont.)

Taking the logarithm of the likelihood simplifies optimization:

$$\log L(\theta) = -\frac{m}{2}\log\frac{1}{\sqrt{2\pi}\sigma} - \frac{1}{2\sigma^2}\sum_{i=1}^{m}\left(y^i - \theta^\top x^i\right)^2$$

Statistics

Minimizing for θ, we get:

$$\min_{\theta} \frac{1}{m}\sum_{i=1}^{m}\left(y^i - \theta^\top x^i\right)^2$$

With Gaussian noise, minimizing MSE equals maximum likelihood estimation

Programming

The ISYE 4600 Colab notebook provides a hands-on review of Python tools for ML, focusing on data handling, modeling, and experimentation throughout the course