AIT 511: Machine Learning – Lecture 1

Motivation & Foundations

Aswin Kannan
IIIT Bangalore

Why Machine Learning? – Deeper Motivation

- Explicit rule-based programming fails for complex perceptual tasks (e.g., Vision, Speech)
- ML enables systems to learn from data patterns, not handcrafted rules
- Emerging data-driven scientific methods complementing theoretical modeling
- Bridging gaps in domains with incomplete knowledge (e.g., Genomics, Financial Forecasting)
- Automation in AI: Adaptability, Scalability, and Continuous Learning from Data

Learning from Nature – Inspirations for ML

- **Visual Cortex Hierarchies → Convolutional Neural Networks (CNNs)**

➢ Humans don’t recognize objects by memorizing pixels. Our visual cortex processes visual signals layer by layer—edges first, then patterns, then object shapes. CNNs mimic this layered processing. They learn basic features (edges, corners) in initial layers, and complex structures (faces, cars) in deeper layers.

➢ **ML Lesson:** Feature hierarchies reduce manual effort in feature engineering and allow scalable learning from raw data.

Languages - Continued

- **Contextual Learning in Language → Transformers (BERT, GPT)**

- Children don’t learn language by rote-learning grammar rules. They grasp words by observing **context**—how words fit in phrases, sentences, situations. Transformer models like BERT and GPT are designed on this idea. They learn contextual embeddings where a word’s meaning depends on its surroundings.

- **ML Lesson:** Models that understand context (not just word counts) handle ambiguity and variability in human language effectively.

Immune System -- Continued..

- **Immune System Adaptability → Anomaly Detection**

- The human immune system doesn’t have a predefined list of threats. It learns to distinguish between ‘self’ and ‘non-self’ by continuously monitoring patterns in the body. Any deviation (like a virus or bacteria) is flagged as an anomaly. ML anomaly detection works similarly—detecting outliers in data streams where normal patterns evolve over time.

- **ML Lesson:** Anomaly detection systems need adaptive learning mechanisms to keep up with changing environments, much like immune responses.

Topics

- First half: Linear Regression, Polynomial Regression, L1, L2 Norm-Based Regularization, Maximum Likelihood Estimate, Bayesian Approach, MAP, Bayes' Classifier, K-NN Classifier, PCA, Decision Tree, Random Forest, and Boosting/Gradient Boosting.

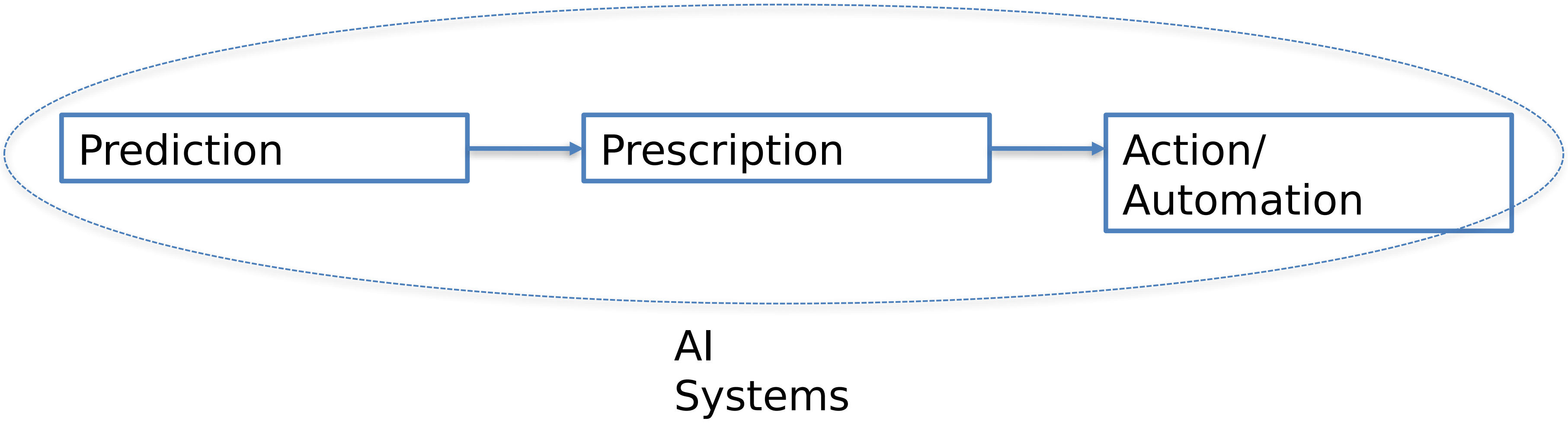
- Second half: Distance Measures, K-Means Clustering, Hierarchical Clustering. Gaussian Mixture Modelling, SVM, Logistic Regression, Neural Networks, Advanced Topics: CNN, and Transformers.

Grading

- Project: 20 points.
- Mid-term: 15 points.
- Attendance: 5 points.
- Quiz: 10 points.

Machine Learning within AI Systems

- AI as a larger system encompassing Reasoning, Planning, Learning, and Acting.
- ML focuses on enabling 'learning' from data streams to improve AI capabilities.
- Supervised, Unsupervised, and Reinforcement Learning paradigms define learning approaches.



Some Sample Data

person_ag	person_ec	person_in	person_er	person_hc	loan_amn	loan_inter	loan_int_r	loan_perc	cb_persor	credit_scc	previous_	loan_statu
22	Master	71948	0	RENT	35000	PERSONAI	16.02	0.49	3	561	No	1
21	High Scho	12282	0	OWN	1000	EDUCATIC	11.14	0.08	2	504	Yes	0
25	High Scho	12438	3	MORTGAC	5500	MEDICAL	12.87	0.44	3	635	No	1
23	Bachelor	79753	0	RENT	35000	MEDICAL	15.23	0.44	2	675	No	1
24	Master	66135	1	RENT	35000	MEDICAL	14.27	0.53	4	586	No	1
21	High Scho	12951	0	OWN	2500	VENTURE	7.14	0.19	2	532	No	1
26	Bachelor	93471	1	RENT	35000	EDUCATIC	12.42	0.37	3	701	No	1
24	High Scho	95550	5	RENT	35000	MEDICAL	11.11	0.37	4	585	No	1
24	Associate	100684	3	RENT	35000	PERSONAI	8.9	0.35	2	544	No	1
21	High Scho	12739	0	OWN	1600	VENTURE	14.74	0.13	3	640	No	1
22	High Scho	102985	0	RENT	35000	VENTURE	10.37	0.34	4	621	No	1

price	sales	SKU
0.333147	4419.991	1
0.517043	4418.936	1
0.7215	4417.645	1
0.892741	4416.46	1
1.538121	4410.973	1
2.993563	4389.891	1
3.174029	4386.073	1
3.53651	4377.312	1
3.777731	4370.568	1
3.977841	4364.346	1
4.109194	4359.924	1
4.546421	4343.024	1
4.602456	4340.591	1
4.68516	4336.88	1

Datasets: Top (Loan lending problem).

Bottom: Retail pricing.

Regression – Problem Formulation

- Objective: Predict continuous outcomes based on input features
- Example: Predicting max temperature from morning temperature readings
- Mathematical Goal: Learn a hypothesis function $y = f(x)$ that generalizes well to unseen data

Machine Learning as Optimization

- Learning = Finding parameters that minimize a chosen loss function
- Common Example: Least Squares Loss for Regression
- Optimization challenges arise with high-dimensional data, noisy observations

Generalization vs Overfitting

- Generalization: Model's ability to perform well on unseen data
- Overfitting: Model captures noise or spurious patterns in training data
- Solution: Regularization techniques to balance model complexity and data fit

Regularization – Intuition

- Adding penalty terms to discourage overly complex models
- L2 Regularization (Ridge): Penalizes large weights, smoothens predictions
- L1 Regularization (LASSO): Encourages sparsity, feature selection

Maximum Likelihood Estimation (MLE)

- MLE: Estimating parameters that maximize the likelihood of observed data
- Linear Regression can be viewed as MLE under Gaussian noise assumption
- Links deterministic models with probabilistic interpretations

Bayesian Approach – Prior Knowledge in Learning

- Bayesian Methods incorporate prior beliefs into learning process
- MAP Estimation: Balancing data likelihood with prior information
- Heard of Bayesian Optimization?

K-Nearest Neighbors (K-NN) Classifier

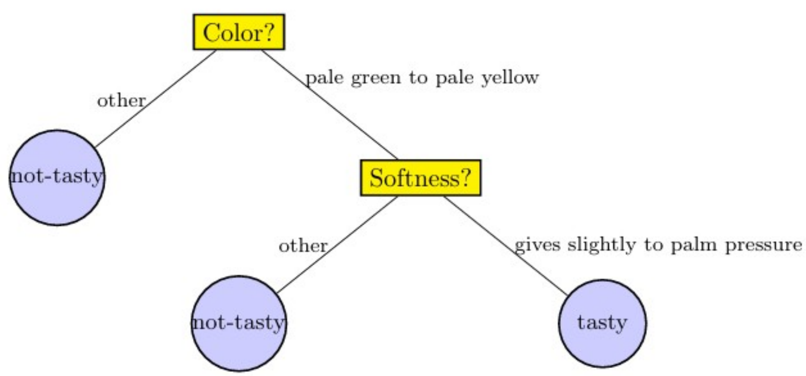
- Instance-based learning without explicit model training
- Predicts based on majority label among K nearest data points
- Simple yet powerful non-parametric method; computationally intensive for large datasets

Principal Component Analysis (PCA)

- Objective: Reduce dimensionality while preserving variance
- Transforms data to new coordinate system of principal components
- Widely used for visualization, noise reduction, and preprocessing

Decision Trees & Ensemble Methods

- Decision Trees: Simple, interpretable models. Operate with IF like conditions. There could be several trees solving our problem!
- Random Forest: Ensemble of Trees to improve generalization via bagging
- Boosting: Sequentially focusing on errors to build strong learners
- Example from 18, page 250).



Summary

- Key Themes: Learning from Data and Generalization.
- We will cover both supervised and unsupervised methods.

Books

- Shai Shalev-Shwartz & Shai Ben-David — *Understanding Machine Learning*
- *Simon Rogers, Introduction to Machine Learning.*