Machine Learning

Guide to AI and Data Courses · Part I: Core

Types of learning | Model and loss | Generalization | Evaluation | Core algorithms | Neighboring courses

Sercan KÜLCÜ · Giresun University · Computer Engineering

© Sercan Külcü. All rights reserved. · Interactive HTML version of these slides: https://sercankulcu.github.io/files/ai_courses/slides/Chapter_03_Machine_Learning.html

What Question Does It Ask?

How does a program generalize from examples?

Machine learning studies algorithms that **learn from examples** without being explicitly programmed and generalize what they learn to new examples. In Mitchell's definition, a program learns if its performance on task T, measured by P, improves with experience E.

At a glance

APPROACH statistical learning	**INPUT** labeled or unlabeled examples	**OUTPUT** a model that makes predictions
SUCCESS MEASURE error on unseen data	**MATH LOAD** high: linear algebra, probability, optimization	**DATA NEEDS** moderate to large

A Short History

1957
Rosenblatt introduces the perceptron, one of the first models to tune its weights from examples.

1959
Samuel's checkers program improves by playing itself; the term "machine learning" spreads.

1986
Quinlan publishes the ID3 decision tree; the same year backpropagation revives neural networks.

1995
Cortes and Vapnik publish support vector machines (SVM).

2001
Breiman introduces random forests: a vote among many trees.

2012
AlexNet wins the ImageNet challenge by a wide margin; the deep learning era begins.

THE COURSE

Core Concepts

Supervised learning
Classification and regression from labeled examples.

Unsupervised learning
Clustering and dimensionality reduction on unlabeled data.

Model and hypothesis
The form and parameters of the function to be learned.

Loss function
Measures how wrong a prediction is; training minimizes it.

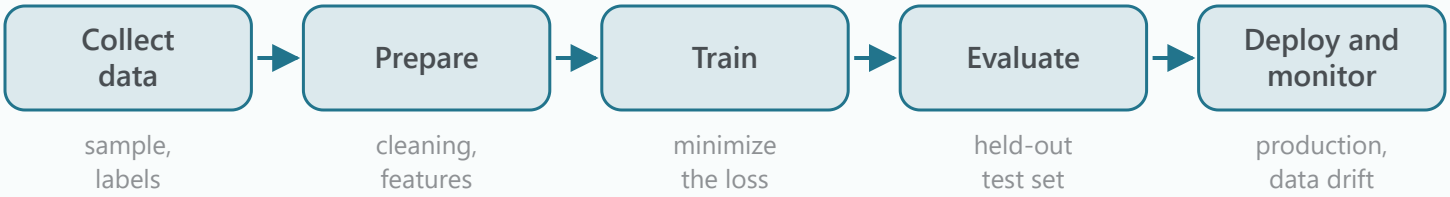
Generalization
Correct predictions on examples not seen in training.

Overfitting
Memorizing the training data; the bias-variance tradeoff.

Validation
Training, validation and test sets; cross-validation.

Features
Turning raw data into numbers a model can use.

A Typical Workflow



Example: A spam filter collects past messages and their labels, turns word counts into features, trains a logistic regression, measures accuracy on held-out test messages and classifies new incoming mail. No rules are written by hand; they are learned from data.

THE COURSE

What Does the Course Cover?

1. The learning problem; supervised and unsupervised learning
2. Linear regression and gradient descent
3. Logistic regression and classification
4. Evaluation: train/test split, cross-validation
5. Overfitting, regularization, bias-variance
6. Decision trees
7. Ensembles: random forests, gradient boosting
8. Support vector machines and kernels
9. k-nearest neighbours and naive Bayes
10. Clustering: k-means, hierarchical clustering
11. Dimensionality reduction: PCA
12. Feature engineering and model selection

Skills you will gain

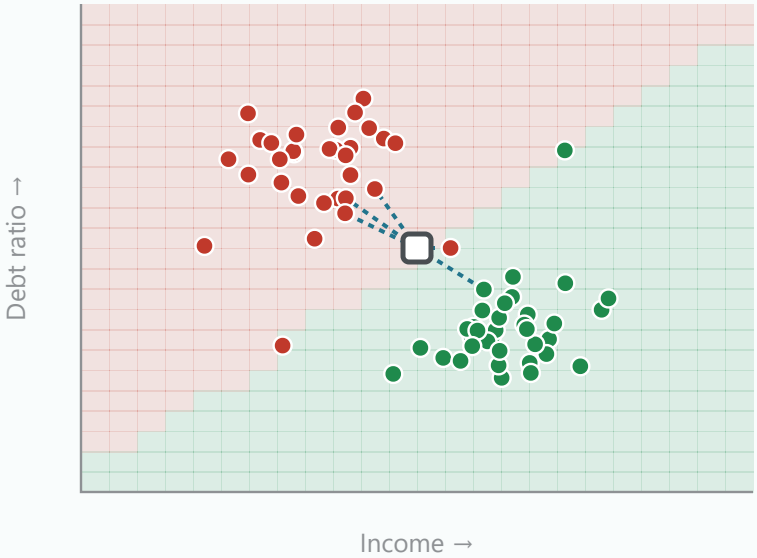
Python and NumPy | data preparation with pandas | scikit-learn | cross-validation | choosing the right metric | reporting results with plots and tables

A typical semester; the order and coverage vary between universities.

TRY IT

Should We Grant the Loan?

The bank knows the income, debt ratio and repayment outcome of past applications. Click the chart to add a new application; the model decides by looking at the **k** most similar past ones.



Neighbors k = **5**

Show decision regions

k = 5: of the 5 nearest applications, 1 repaid and 4 defaulted.
Decision: **decline the loan** (majority vote).

● repaid ● defaulted. With k = 1, see how the regions split around single examples: that is **overfitting**.

Nobody wrote the rule: the decision comes entirely from **examples**. With a small k the model memorizes noise, with a large k it misses detail; the right k is chosen on held-out test data.

CONTEXT

Prerequisites

Courses worth knowing before this one, and why they are needed.

Course	Why it is needed
Artificial Intelligence	an agent that learns from data; statistical generalization instead of search
Linear Algebra	feature vectors, matrix multiplication, PCA
Probability and Statistics	generalization error, bias-variance, probabilistic models
Optimization	loss minimization, regularization
Programming	reading data, loops and functions in Python or Java
Calculus	derivatives, partial derivatives and the chain rule: the language of gradient descent

CONTEXT

Are You Ready for This Course?

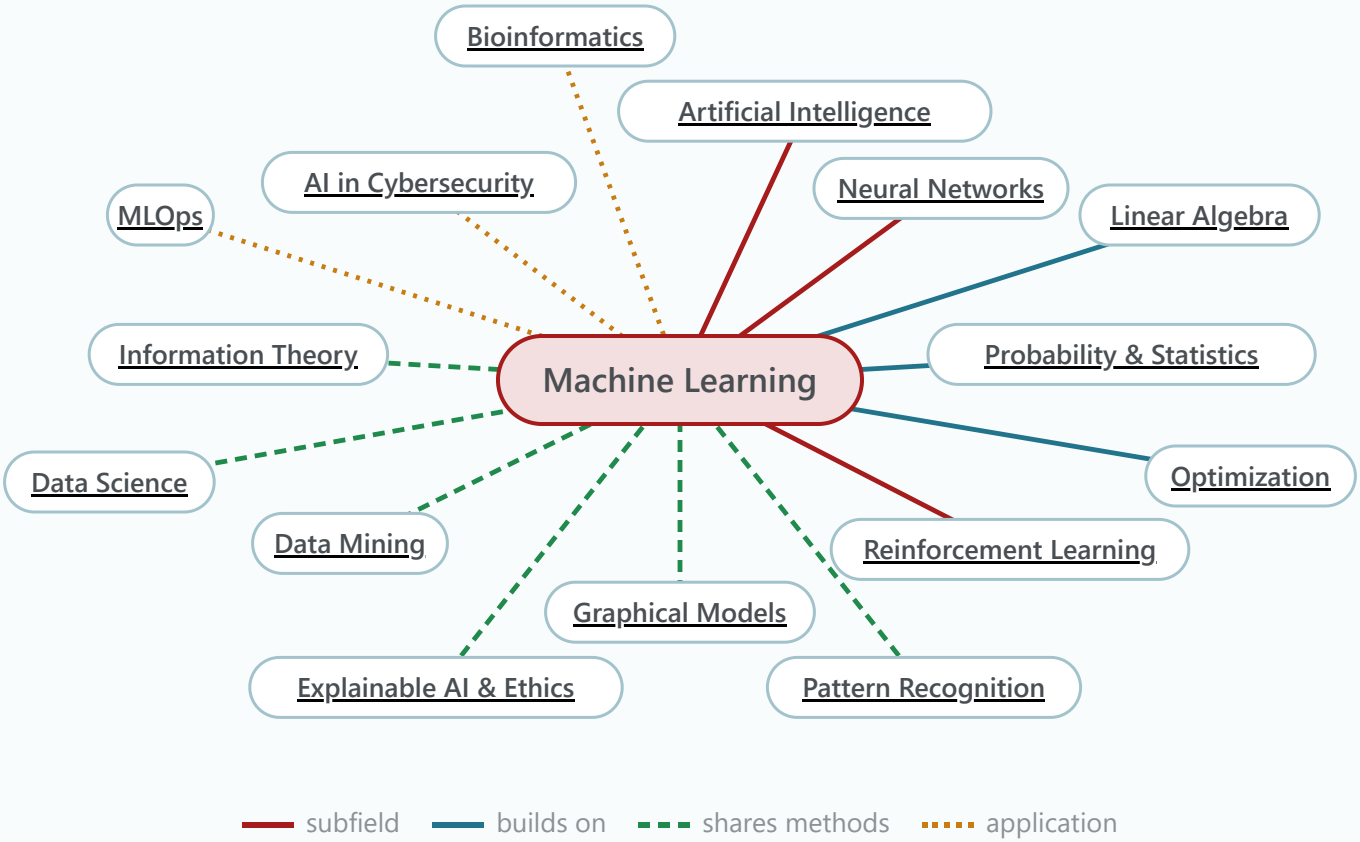
Ask yourself; for every "no", the right-hand column shows where to start.

Ask yourself	If not, start with
Can you take a derivative and a partial derivative?	Derivatives and the chain rule from calculus
Can you multiply matrices and take a dot product?	Linear Algebra
Can you explain mean, variance and conditional probability?	Probability and Statistics
Can you write loops and functions in Python and read a CSV file?	An intro programming course, then a short pandas tutorial

If you can say "yes" to most of these, you are ready; one or two gaps can be closed during the semester.

CONTEXT

Neighbouring Courses



Overlapping Courses (1/2)

Relation	Course	Shared ground
subfield	This course is a subfield of Artificial Intelligence	an agent that learns from data; statistical generalization instead of search
subfield	Artificial Neural Networks is a subfield of this course	parametric model, loss function, gradient descent
subfield	Reinforcement Learning is a subfield of this course	learning from rewards instead of labels
builds on	This course builds on Linear Algebra	feature vectors, matrix multiplication, PCA
builds on	This course builds on Probability and Statistics	generalization error, bias-variance, probabilistic models
builds on	This course builds on Optimization	loss minimization, regularization
application	Machine Learning Systems and MLOps applies the methods of this course	deploying, monitoring and retraining models
application	AI in Cybersecurity applies the methods of this course	intrusion detection, malware classification

Overlapping Courses (2/2)

Relation	Course	Shared ground
application	Bioinformatics applies the methods of this course	gene expression classification, protein structure
shares methods	Shares methods with Pattern Recognition	classifiers, decision boundaries
shares methods	Shares methods with Probabilistic Graphical Models	probabilistic classifiers, EM algorithm
shares methods	Shares methods with Explainable AI and AI Ethics	model explanation, feature importance
shares methods	Shares methods with Data Mining	same algorithms; the goal is discovery, not prediction
shares methods	Shares methods with Data Science	predictive models
shares methods	Shares methods with Information Theory	cross-entropy loss, decision tree split criterion

DIFFERENCES

Where It Differs (1/2)

Criterion	Machine Learning	Artificial Intelligence	Data Mining
Core question	How does a program generalize from examples?	How does a machine act intelligently by searching, reasoning and planning?	How are hidden, useful patterns found in large piles of data?
Approach	statistical learning	search, logic, probability and learning combined	discovery-driven analysis
Input	labeled or unlabeled examples	problem definition: states, actions, goal	large tables or transaction data
Output	a model that makes predictions	action sequence, plan or decision	rules, clusters, outliers
Success measure	error on unseen data	reaching the goal, cost, rationality	interestingness, support, confidence
Math load	high: linear algebra, probability, optimization	medium: logic, probability, graphs	medium
Data needs	moderate to large	little; often a model and knowledge suffice	large

- **Artificial Intelligence:** AI is the framework; ML focuses only on learning from data within it.
- **Data Mining:** They share algorithms; ML aims to predict new examples, data mining to discover patterns in data.

DIFFERENCES

Where It Differs (2/2)

Criterion	Machine Learning	Probability & Statistics	Deep Learning
Core question	How does a program generalize from examples?	How is uncertainty measured, and how do we draw reliable conclusions from limited data?	How do many-layered networks learn useful representations from raw data?
Approach	statistical learning	probability models and inference	multilayer representation learning
Input	labeled or unlabeled examples	sample data	raw data: images, text, audio
Output	a model that makes predictions	estimate, confidence interval, test result	learned representation and prediction
Success measure	error on unseen data	unbiasedness, power, significance	accuracy, loss, scaling
Math load	high: linear algebra, probability, optimization	the theory itself	high: derivatives, matrices, probability
Data needs	moderate to large	a sample	large; needs GPUs

- **Probability & Statistics:** Statistics emphasizes a model's meaning and uncertainty, ML its predictive performance.
- **Deep Learning:** Deep learning is a branch of ML; it learns features from raw data instead of hand-designing them.

PRACTICE

Applications

Spam filtering
Classifying emails by their content.

Credit scoring
Predicting an applicant's risk of default.

Recommendation
Products or content a user is likely to enjoy.

Medical prediction
Estimating disease risk from patient data.

Demand forecasting
Sales and inventory planning.

Fraud detection
Catching unusual card transactions.

PRACTICE

Risks and Ethics

Biased data, biased decisions

Discrimination in historical data passes straight into the model: hiring or credit models can systematically exclude some groups.

Data leakage

When test information leaks into training, a model looks flawless on paper and fails in the field.

Privacy

Personal information in the training data can be extracted from the model; GDPR and similar laws limit how data may be used.

Misplaced confidence

High accuracy collapses when the data distribution shifts, and the model looks just as confident on cases it does not know.

In depth: Explainable AI and AI Ethics.

PRACTICE

Tools and Datasets

Tools and libraries

scikit-learn | pandas | NumPy | XGBoost | Weka | R

Well-known datasets

Iris | MNIST | UCI Machine Learning Repository | Titanic (Kaggle) | Adult (income) | Wine

PRACTICE

Common Misconceptions

Common belief	In fact
A more complex model is always better.	A complex model can memorize the training data (overfitting) and do worse on new data.
Training accuracy shows success.	Success is measured on test data never used in training.
More data solves every problem.	More biased or noisy data only multiplies the errors.
The model learns cause and effect.	ML finds correlations; causation needs experiments or causal inference.

NEXT STEPS

After This Course

Courses that build on or extend this one

Artificial Neural Networks | Reinforcement Learning | Machine Learning Systems and MLOps | AI in Cybersecurity | Bioinformatics

Related careers

Machine learning engineer | Data scientist | Researcher | MLOps engineer

Machine learning is the crossroads of the map: neural networks and reinforcement learning are its branches; MLOps, cybersecurity and bioinformatics apply its methods. After this course you can go deeper (neural networks, deep learning) or move on to an application area.

NEXT STEPS

Further Reading

- Mitchell, *Machine Learning*, McGraw-Hill, 1997.
- Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.
- Hastie, Tibshirani and Friedman, *The Elements of Statistical Learning*, 2nd ed., Springer, 2009.
- Alpaydın, *Introduction to Machine Learning*, 4th ed., MIT Press, 2020.
- Géron, *Hands-On Machine Learning with Scikit-Learn, Keras and TensorFlow*, 3rd ed., O'Reilly, 2022.

The End

Chapter 3: Machine Learning

← Previous chapter · Map of the field · Next chapter →