



Overview of Machine Learning

What is Machine Learning?

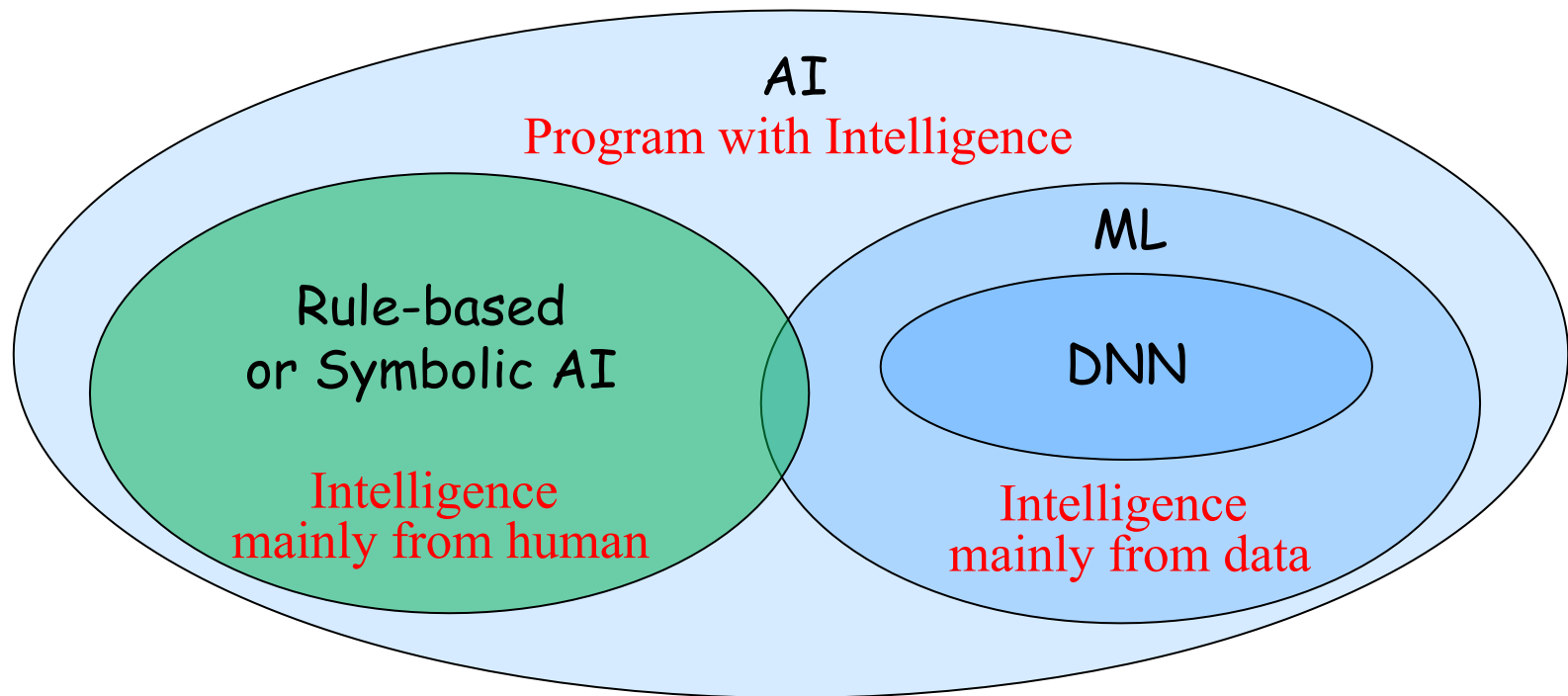
- **Techniques to improve the performance of programs based on given data, previous results, or experiences.**
 - Methods for creating computer programs by the analysis of data sets.
- **Hmm, It sounds like data-analysis techniques**
 - It's not perfect, but it's pretty accurate
- **Then, How is it different from with statistics?**
 - Statistics is a branch of mathematics, but ML is not
 - You can add heuristic knowledge or bold assumptions for data analysis

Applications

- **Voice/Face/Fingerprint/Iris/DNA/Signature recognition**
- **Web-search, Document & information retrieval, Machine translation**
- **Recommendation, Spam filter**
- **Credit card fraud detection, Loan application analysis**
- **Marketing, Stock market prediction**
- **Games: Chess, Go**
- **...**

How they are different

- Artificial Intelligence, Machine Learning, and Deep Neural Network



How they are different

- **Artificial Intelligence**

- To build programs acting intelligently
- We need to code “intelligence” into programs
- Where do we obtain “intelligence”?

- **From human-beings**

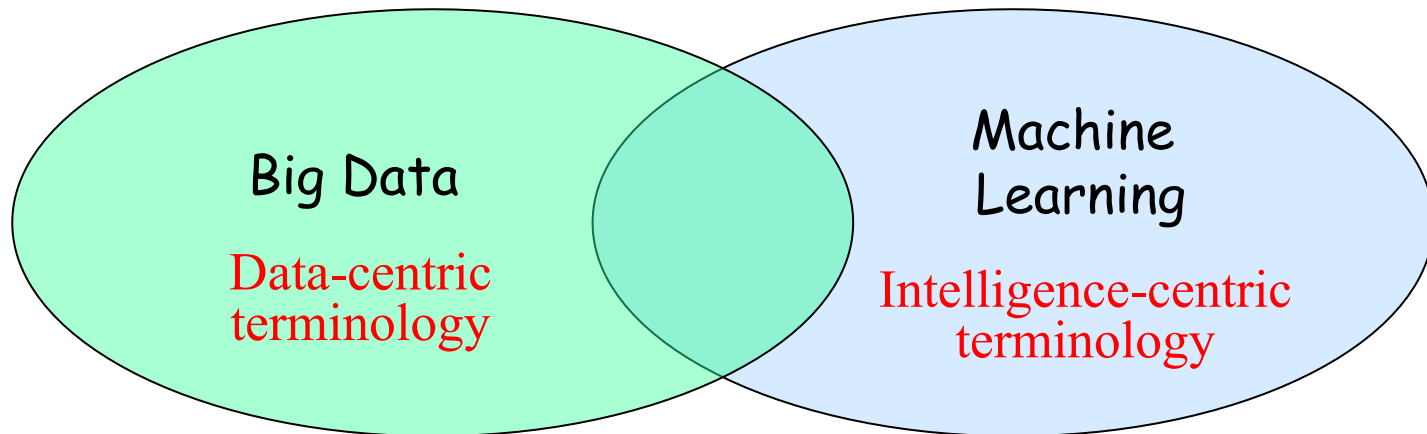
- They have intelligence
- Extract “intelligence” from them by asking “how do you do that?”
- Expert systems -> Big failure

- **From data**

- Human-beings create data, so their intelligent behavior or decisions are resolved in the data
- If we distillate it, we can obtain “intelligence”
- Successful, but why now?

How they are different

- **Big Data and Artificial Intelligence**



How to collect data

How to efficiently manage data

How to efficiently store data

How to efficiently analyze

How to obtain useful patterns from data

How to build intelligent program

Related Theories

■ Math

- Probability and Statistics
 - Conditional Probability, Bayesian Rule, Independency, Maximum Likelihood, Normal Distribution, Expectation, Variance, Covariance, Covariance Matrix
- Linear Algebra
 - Inverse Matrix, Eigen Value, Eigen Vector, Singular Value Decomposition
- Optimization:
 - Gradient Descent, Convex Optimization, Expectation-Maximization, Quadratic Programming

■ Programming

- Dynamic Programming
- Divide and Conquer
- Computational complexity theory

Some Types of ML Problems (1)

- **Classification**

- Voice/Face/Fingerprint/Iris/DNA/Signature recognition, Recommendation, Spam filter, Credit card fraud detection

- **Regression**

- Loan application analysis, Marketing, Stock market prediction

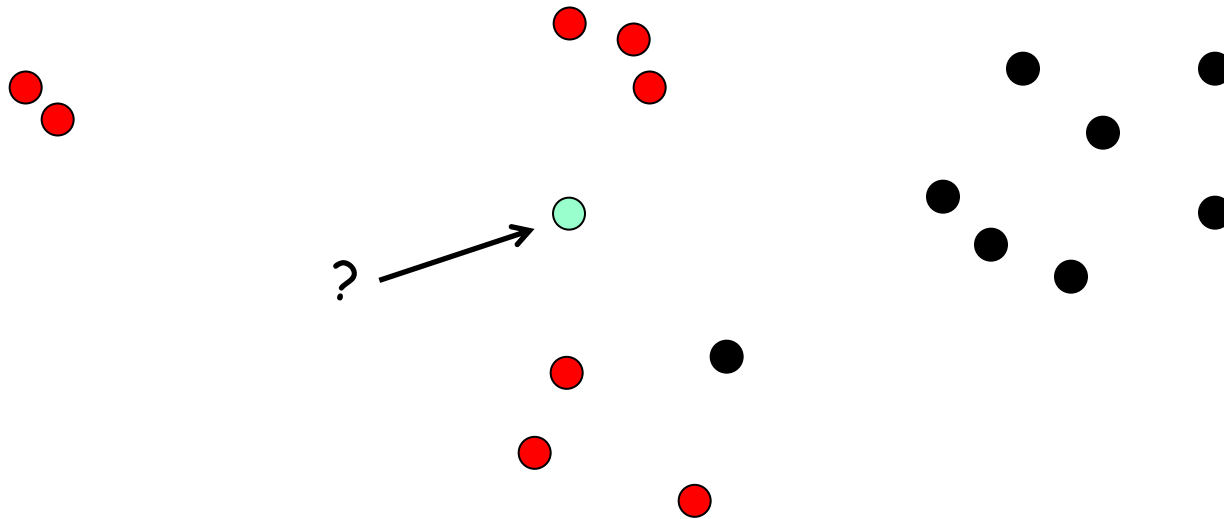
- **Clustering**

- Web-search, Document & information retrieval, Machine translation

Some Types of ML Problems (2)

■ Classification

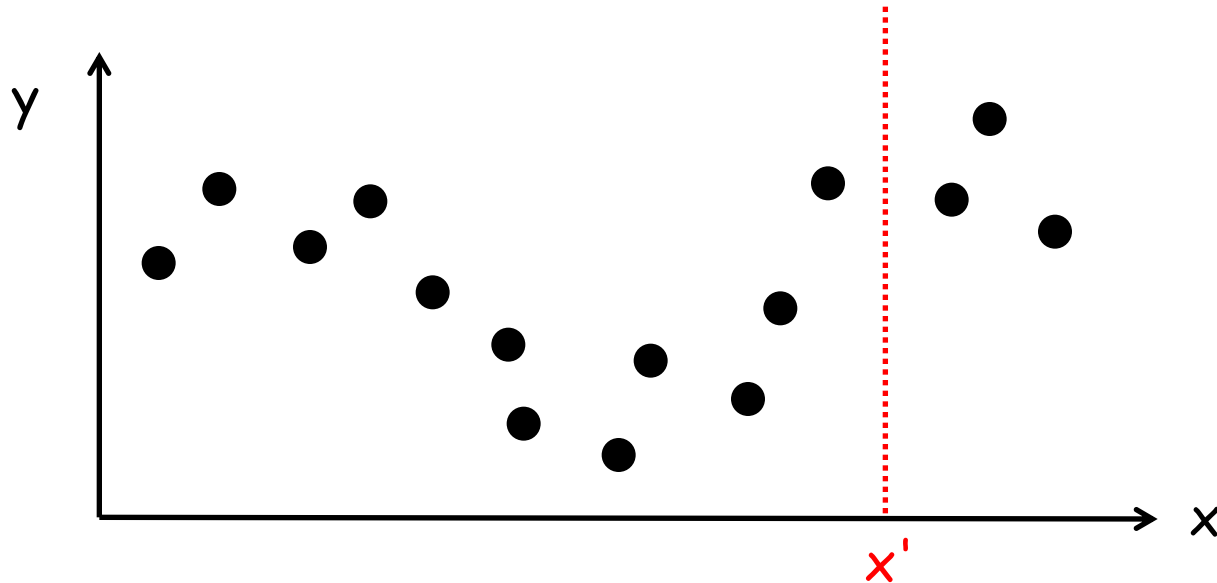
- A set of labeled data is given
- Your program should find the boundary between labels
- If a query is given, your program should answer the label



Some Types of ML Problems (3)

■ Regression

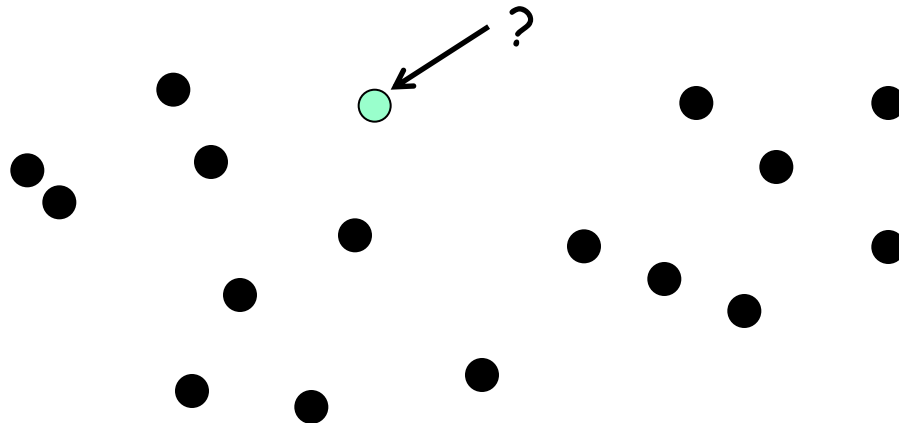
- A set of (\mathbf{x}, y) 's is given. (\mathbf{x} is a vector, y is a real number)
- Your program should find the functional relation between \mathbf{x} and y
- If a query, \mathbf{x}' , is given, your program should answer y for \mathbf{x}'



Some Types of ML Problems (4)

■ Clustering

- Unlabeled data is given.
- Your program should group the data (Finding hidden structure of data)
- If a query is given, your program should determine the group in which the query belongs to



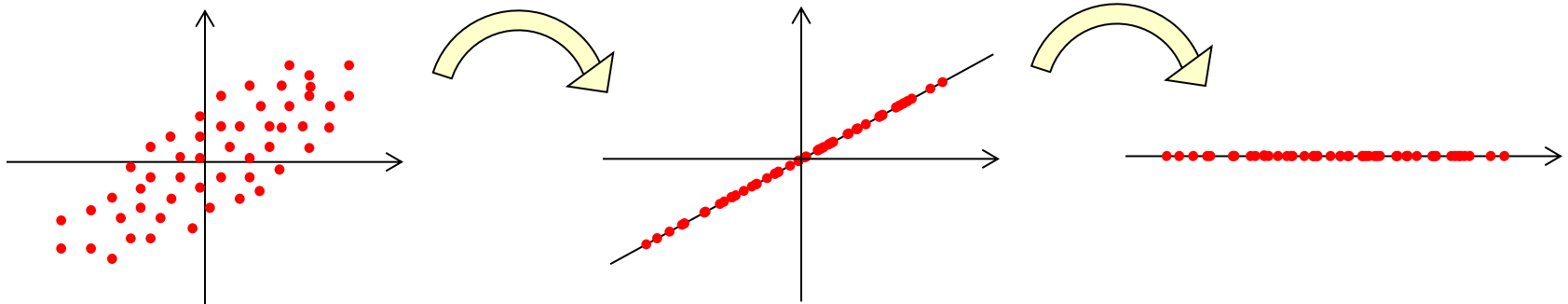
Some Types of ML Problems (5)

■ Dimension Reduction

- A set of unlabeled data is given.
- Your program should reduce the dimension of data by minimizing the loss of information

Find the projection line

Remove other axes



Supervised & Unsupervised Learning

- **Supervised Learning**

- Classification, Regression
- All given data are labeled

- **Unsupervised Learning**

- Clustering, Dimension Reduction, Association
- Data is not labeled