Machine Learning

Introduction - Lecture I

Organization

- Lecture

  Dr. Anastassia Küstenmacher

  anastassia.kuestenmacher@h-brs.de

- Lab-classes

  Ganesamanian Kolappan

- Course webpage (LEA)

Organization

- Structure: weekly 1Lecture + 1Exercise
- Place and Time: every Monday 9:00AM-10:30AM
- Exam: electronic

Organization

Exercises

- Typically 1 exercise assignment every week for the first week.
- Mostly Python based exercises.
- Hands-on exercises with the algorithms from the lecture

Project

- Industry-oriented project: unsupervised learning, data engineering
- Real dataset
- Delivery is a report

Teams are encouraged

- You can form teams of up to 2 people for the exercises
- Each team submits one solution with the list of all team members in the submission (submission to LEA)

Organization

1	12.04.21	Introduction
2	19.04.21	Parametric Methods, Gaussian Distribution, Maximum Likelihood
3	26.04.21	Nonparametric Methods, Histograms, Kernel Density Estimation, Bayesian Learning
4	03.05.21	Mixture of Gaussians, k-Means Clustering, EM-Clustering, EM Algorithm
5	10.05.21	Linear Discriminants I
6	17.05.21	Linear Discriminants II
7	31.05.21	Ensemble Methods and Boosting
8	07.06.21	Randomized Decision Trees, Random Forests, Extremely Randomized Trees, Ferns
9	14.06.21	Reinforcement learning
10	21.06.21	Single-Layer Perceptron, Multi-Layer Perceptron, Mapping to Linear Discriminants, Error Functions, Regularization, Multi-layer Networks, Chain rule, Gradient Descent
11	28.06.21	Backpropagation, Computational Graphs, Stochastic Gradient Descent, Minibatch Learning, Optimizers (Momentum, RMS-Prop, AdaGrad, Adam)
12	05.07.21	Initialization (Glorot, He), Nonliniearities, Drop-out, Batch Normalization, Learning Rate Schedules

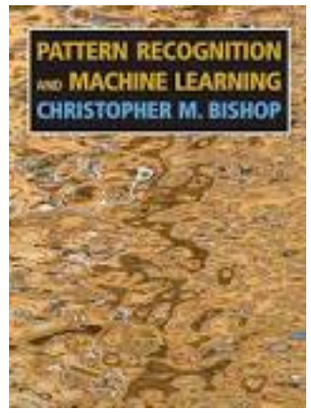
Course Rules

Course Rules

Rules

- Lecture period is 02.11.2020-29.01.2021.
- Lectures take place on Monday start at 9:00AM and will have open end
- Grades are based on homework exercises as well as a final exam.
- The final exam will be in electronic form (written or oral depends on the situation).
- You may choose to have 40% of your final grade calculated by taking your homework assignments into consideration or to simply have the exam count for 100% of your grade i.e. [min (exam, (0,4 * exercises + 0,6 * exam))]. Furthermore in the final exam you need at leat 20% of the points at to pass the course as a whole. Not obtaining this will also anull your homework grade and will make you flunk the course.
- Collaboration within groups of three students or less is OK but *ALL* the students must be able to present *ALL* the exercises.
- You are required to submit: 50% of assignments. Of these, at least 50% must be correct.
- Submit always a PDF-file containing your Ipython notebook.
- Use the following naming convention for your submissions: LA_FirstnameLastname_dateOfLecture.pdf
- If you want to commit EXECUTABLE code to me (that is recommended) then pack everything (i.e.PDF plus code *.m fiels) in one ZIP archive using the same naming convention like e.g.:
    - LA_JohnDoe_YYMMDD.pdf
    - OR
    - LA_JohnDoe_200426.zip
- Note that the date is the date the exercises were handed out to you and not their due date.
- Scan (there is a scanner available in the RoboCup Lab in room C-069) or photograph your handwritten documents to submit them electronically. You may wish to do this in black and white (300dpi is sufficient) as PDF documents larger than 1 Mb will *NOT* be accepted.
- Be certain to have your own printout which you can use to explain your solution on the board in front of the class.
- Any complains and comments about the homework grades are accepted during the week after getting the feedback from the teacher.
- You must deliver exercises through LEA by Saturday 20:00 (sharp!).
- In the case that there are 3 students or less, the lecture will be canceled.

Sources



- Textbooks

C. M. Bishop Pattern Recognition and Machine Learning Springer, 2006

R.O. Duda, P.E. Hart, D.G. Stork Pattern Classification 2nd Ed., Wiley-Interscience, 2000

T. Mitchell, Machine Learning, McGraw Hill, 1997

- Research papers will be given out for some topics.

Material

Slides are adopted from

- Prof. Bastian Leibe RWTH Aachen
- Prof. Gerhard Kraetzschmar

Beginning of Video 1.1

Lecture 1.1

Motivation and Application Examples

Computer vision

Information retrieval

Financial Prediction

Medical Diagnosis

Autonomous driving

etc …

Computer Vision

Object Recognition

Segmentation

Scene Understanding

Slide source: B. Leibe Slides ML, 2017

Information retrieval

machine learning

All Images News Videos Books More Settings Tools

About 334.000.000 results (0,75 seconds)

Machine Learning | Ihr Partner für Innovationen
Ad www.maschinendaten-analyse.de/
Maschinendaten intelligent nutzen - Wir bieten Ihnen individuelle Lösungen!
Über 25 Jahre Expertise · Bundesweit im Einsatz · Über 300 IT Experten · Kostenlose Beratung

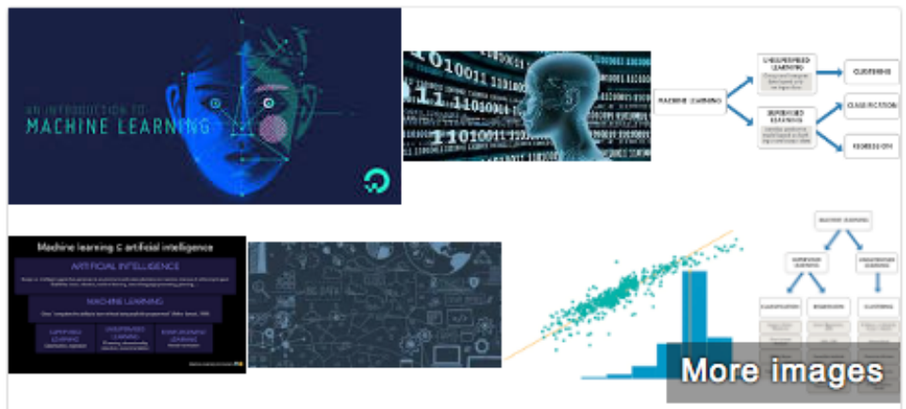
Machine Learning | Free Best Practices Guide | SAS.com
Ad www.sas.com/
Statistics and **Machine Learning** at Scale. Get Your Copy of the Guide Now!
Business Analytics · Predictive Analytics · Statistical Analytics
Highlights: Operating Since 1976, Offers Data Management Software And Services
Buy SAS Analytics · Analytics White Papers · All Analytics Products · Global Analytics Webinars

Machine Learning Plattformen | Skalierbare Systeme erzeugen
Ad de.mathworks.com/machinelearning/plattform
Erkenntnisse eines **Machine Learning** & Data Mining Wettbewerbs.
Neural Networks Erstellen · MATLAB 30 Tage Testen · Deep Learning mit MATLAB

C3 IoT - Platform for AI & IoT | Machine Learning Applications | c3iot.com
Ad www.c3iot.com/Enterprise-AI/MachineLearning
Comprehensive and Proven Platform for Enterprise Digital Transformation
Products: C3 IoT Platform, C3 IoT Applications, C3 Enterprise Data Lake
Contact Us · C3 IoT - Company Overview · Customers · Customer Videos · C3 Type System

Machine learning - Wikipedia
https://en.wikipedia.org/wiki/Machine_learning
Machine learning is a field of computer science that gives computer systems the ability to "learn" (i.e. progressively improve performance on a specific task) with data, without being explicitly programmed.

https://de.mathworks.com/campaigns/products/ppc/google/scalable-machine-learning-system.html



Machine learning

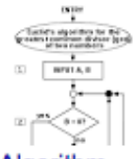
Field of study

Machine learning is a field of computer science that gives computer systems the ability to "learn" with data, without being explicitly programmed. The name Machine learning was coined in 1959 by Arthur Samuel. Wikipedia

People also search for

View 3+ more


Artificial intelligence


Algorithm


Deep learning

Big data

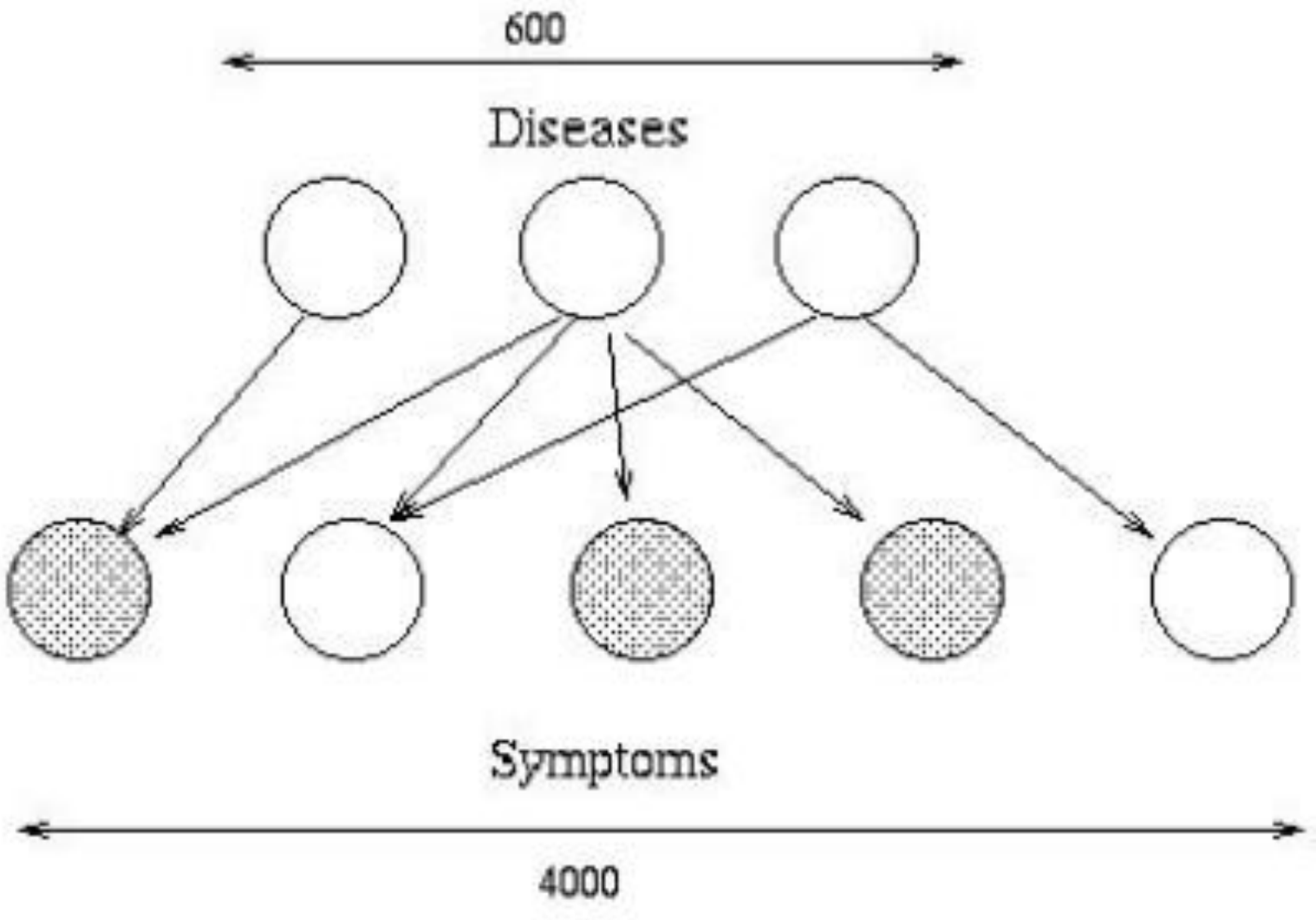
Natural-la… processing

Financial Prediction

Time series analysis

Slide source: B. Leibe Slides ML, 2017

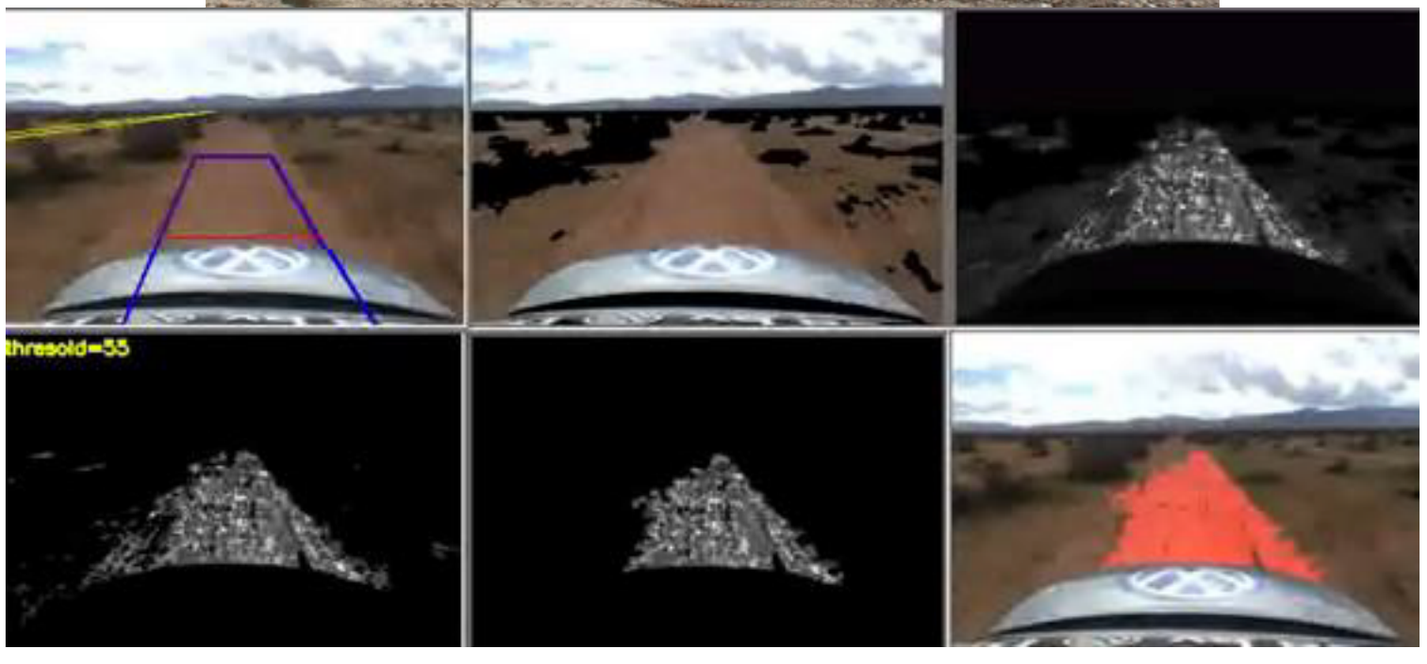
Medical Diagnosis



Inference from partial observations

Slide source: B. Leibe Slides ML, 2017

Autonomous Driving



Slide source: B. Leibe Slides ML, 2017

Motivation

Statistical Machine Learning

- Principles, methods and algorithms for learning and prediction on the basis of past evidence

Already everywhere

- Speech recognition (e.g. speed-dealing)
- Computer vision (e.g. face detection)
- Hand-written character recognition (e.g. letter delivery)
- Information retrieval (e.g. image & video indexing)
- Operation systems (e.g. caching)
- Fraud detection (e.g. credit cards)
- Text filtering (e.g. email spam filters)
- Game playing (e.g. strategy prediction)
- Robotics (e.g. prediction of battery lifetime)

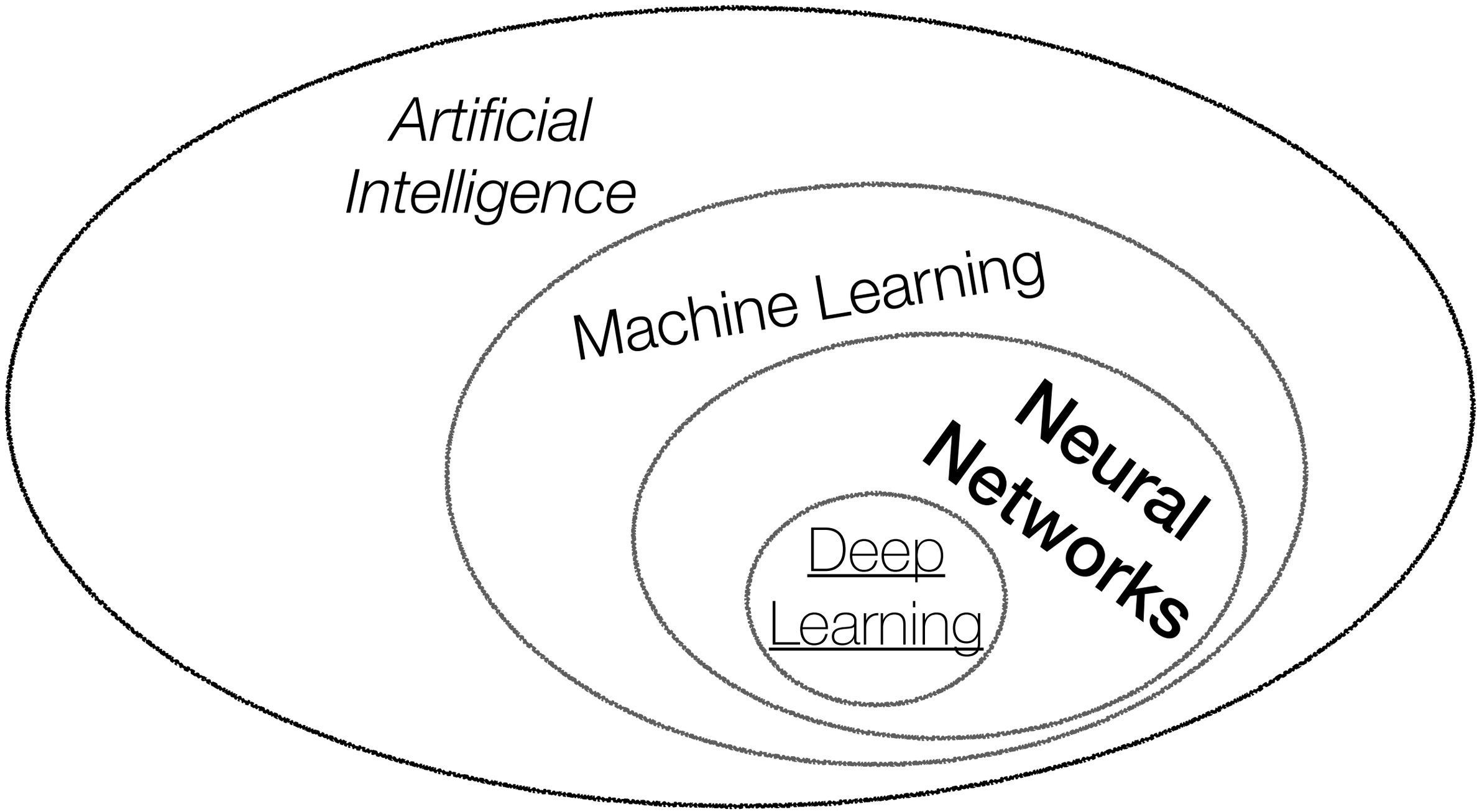
Slide source: B. Leibe Slides ML, 2015

End of Video 1.1

Introduction to Machine Learning

Lecture 1.2

- What is “learn"?
- What is “perform"?
- What is “task"?
- What is “experience“?


Artificial
Intelligence
Machine Learning
Neural
Networks
Deep
Learning

Why Machine Learning?

Goal

- Machines that **learn** to **perform** a **task** from **experience**

Reason

- Crucial component of every intelligent/autonomous systems
- Important for a system's adaptability
- Important for a system's generalisation capabilities
- Attempt to understand human learning

Slide source: B. Leibe Slides ML, 2015

What is Machine Learning?

Learning to perform a task from experience

- Most important part here!
- We do not want to encode the knowledge ourselves.
- The machine should **learn** the relevant criteria automatically from past observations and **adapt** to the given situation.

Tools:

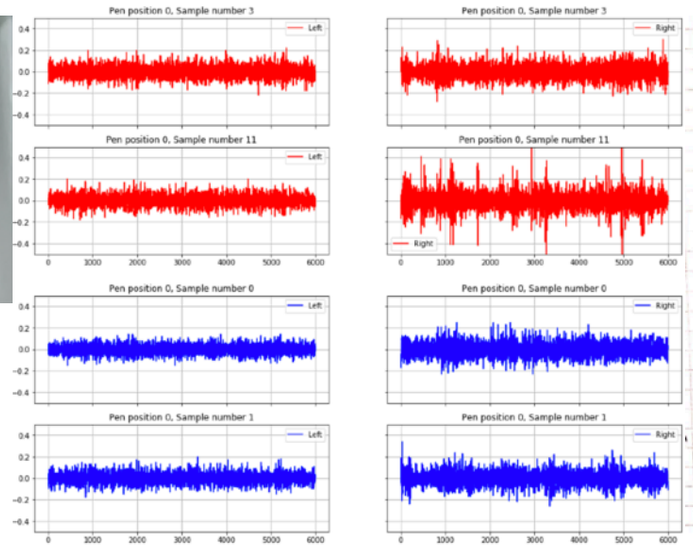
- Statistics, Probability theory, Decision theory, Information theory, Optimization theory

Slide source: B. Leibe Slides ML, 2015

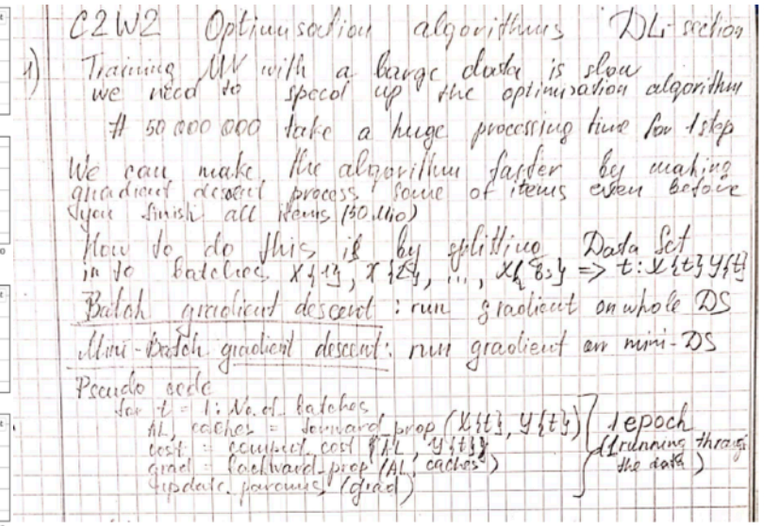
What is Machine Learning?

Learning to perform a **task** from experience

Classify: Good or bad



Recognise text



Predict

Can often be expressed through a mathematical function

$$\textit{Output} \longrightarrow y = f(x, w) \longleftarrow \textit{Parameters}$$

$$\textit{Input} \nearrow$$

(what is "learned")

Slide source: B. Leibe Slides ML, 2015

What is Machine Learning?

Learning to **perform** a task from experience

Performance: "99% correct classification"

- Of what???
  - Characters? Words? Sentences?
  - Speaker/writer independent?
  - Over what data set?
  - …

"The car drives without human intervention 99% of the time on country roads"

Image source: wikimedia

Slide source: B. Leibe Slides ML, 2015

What is Machine Learning?

Learning to **perform** a task from experience

Performance measure: Typically one number

- % correctly classified letters
- Average driving distance (until crash…)
- % games won
- % correctly recognised words, sentences, answers

Generalization performance

- Training vs. test
- “All” data

Slide source: B. Leibe Slides ML, 2015

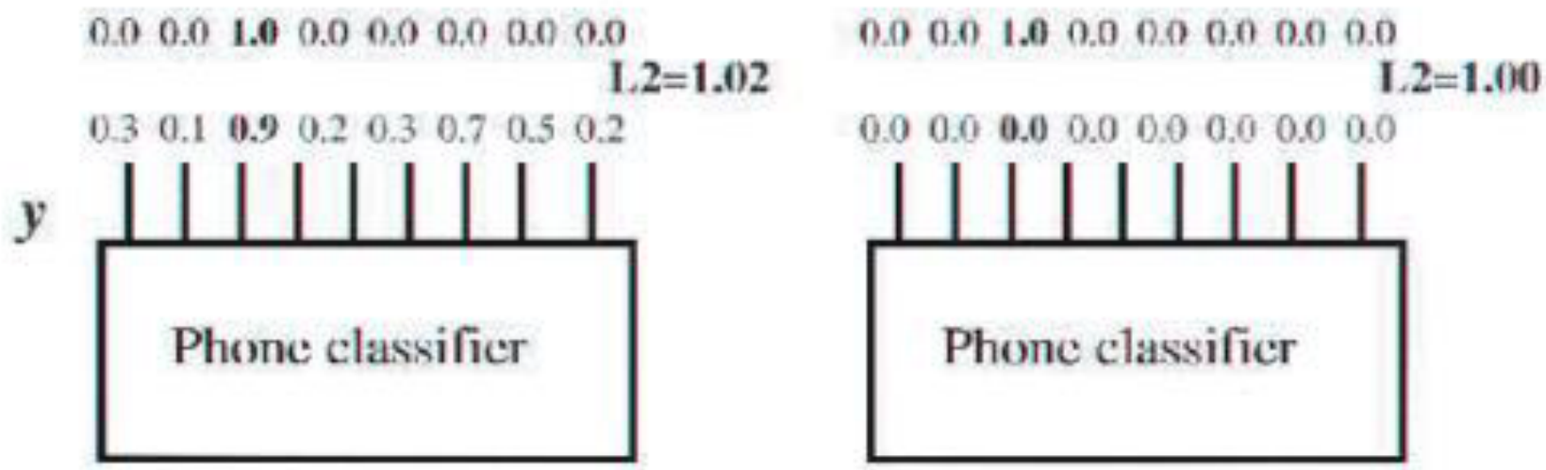
What is Machine Learning?

Learning to **perform** a task from experience

Performance measure: more subtle problem

- Also necessary to compare partially correct outputs.
- How do we weight different kinds of errors?

- Example: L2 norm



Slide source: B. Leibe Slides ML, 2015

What is Machine Learning?

Learning to perform a task from **experience**

What data is given?

Data with labels: **supervised learning**

- Images / speech with target labels
- Car sensor data with target steering signal

Data without labels: **unsupervised learning**

- Automatic clustering of sounds and phonemes
- Automatic clustering of web sites

Some data with, some without labels: **semi-supervised learning**

No examples: **learning by doing**

Feedback/rewards: **reinforcement learning**

Slide source: B. Leibe Slides ML, 2015

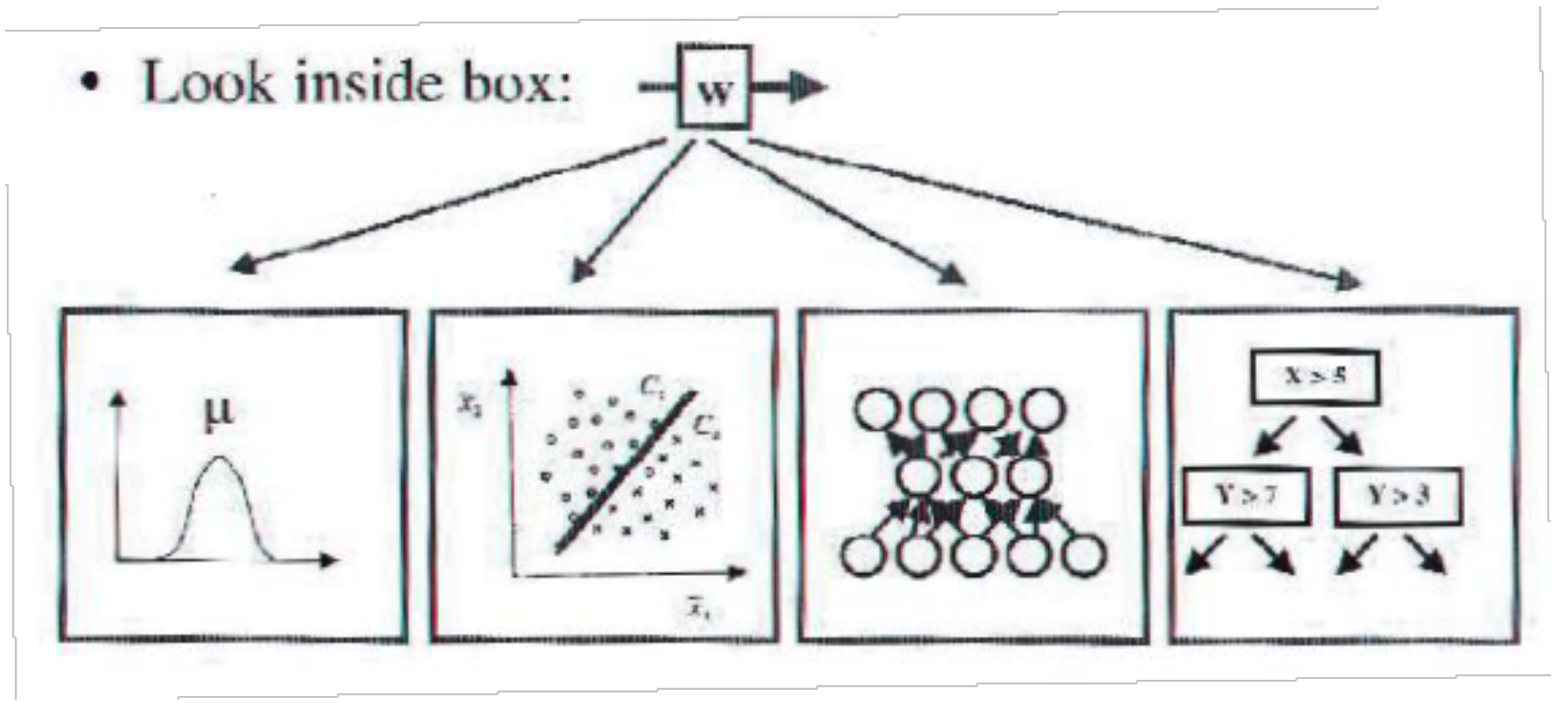
What is Machine Learning?

$$y = f(x; w)$$

w: characterises the family of functions

w: indexes the space of hypotheses

w: vector, connection matrix, graph, …

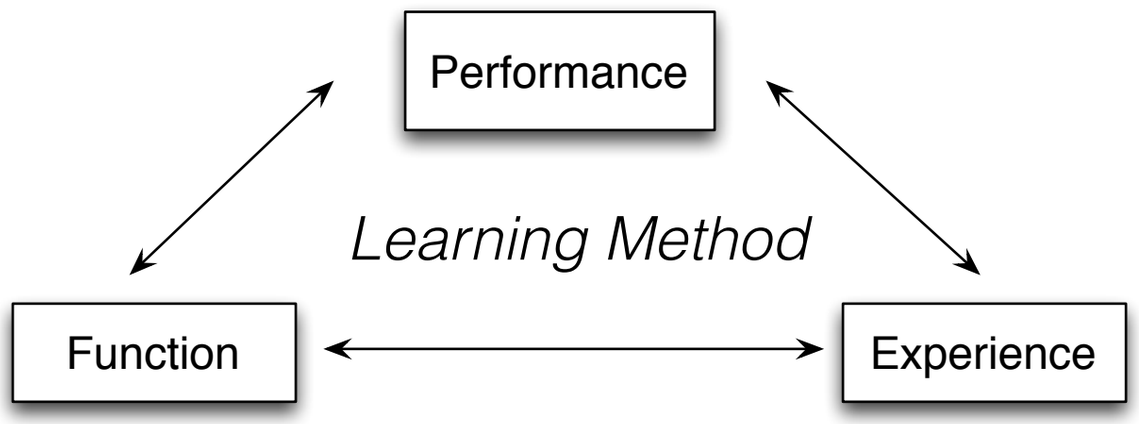


Slide source: B. Leibe Slides ML, 2015

What is Machine Learning?

Learning to perform a task from experience

- Most often learning = optimization
- Search in hypothesis space
  - Search for the “best” function / model parameter w
    - I.e. maximize $y = f(x; w)$ w.r.t. the performance measure



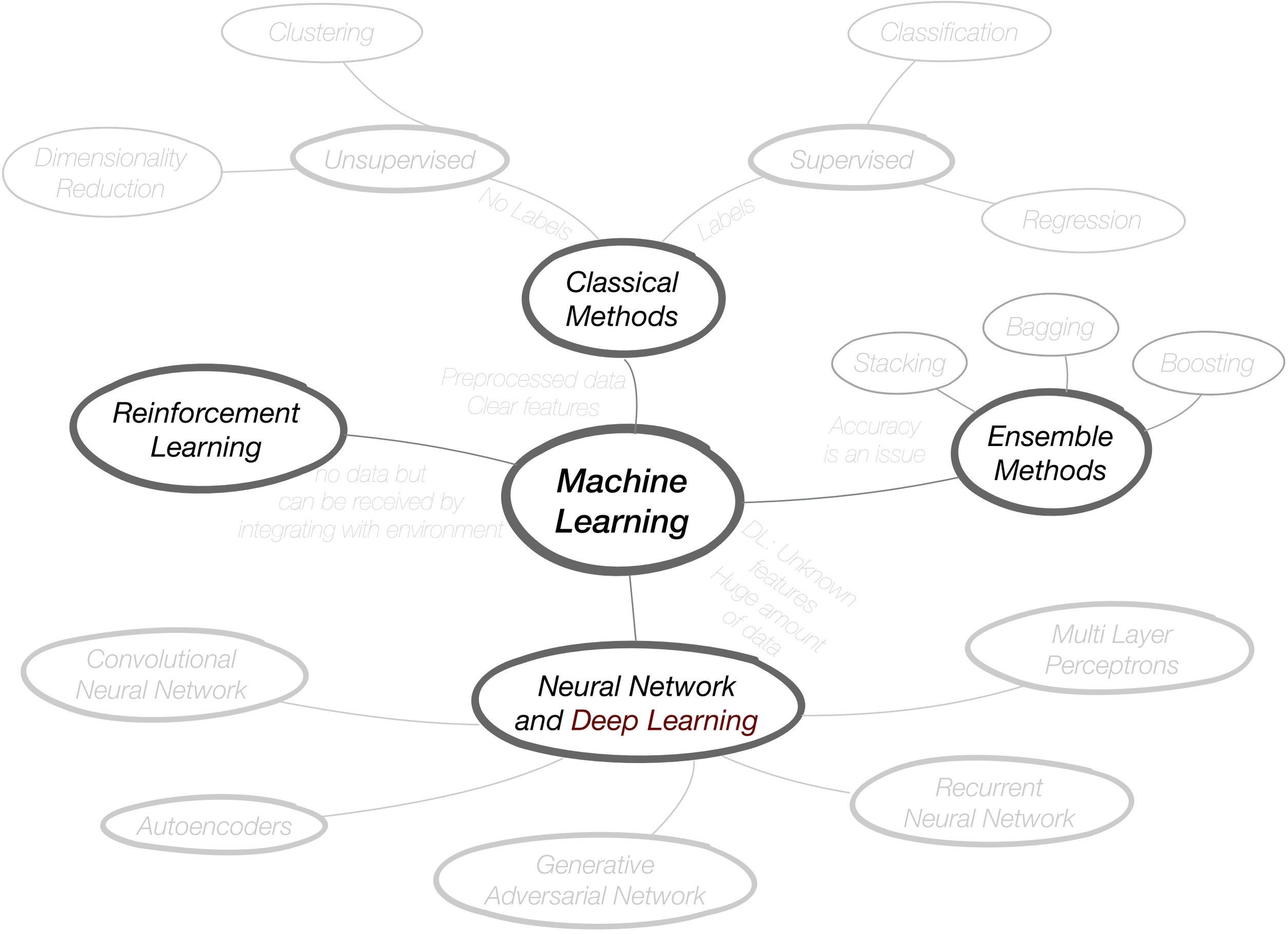
Slide source: B. Leibe Slides ML, 2015

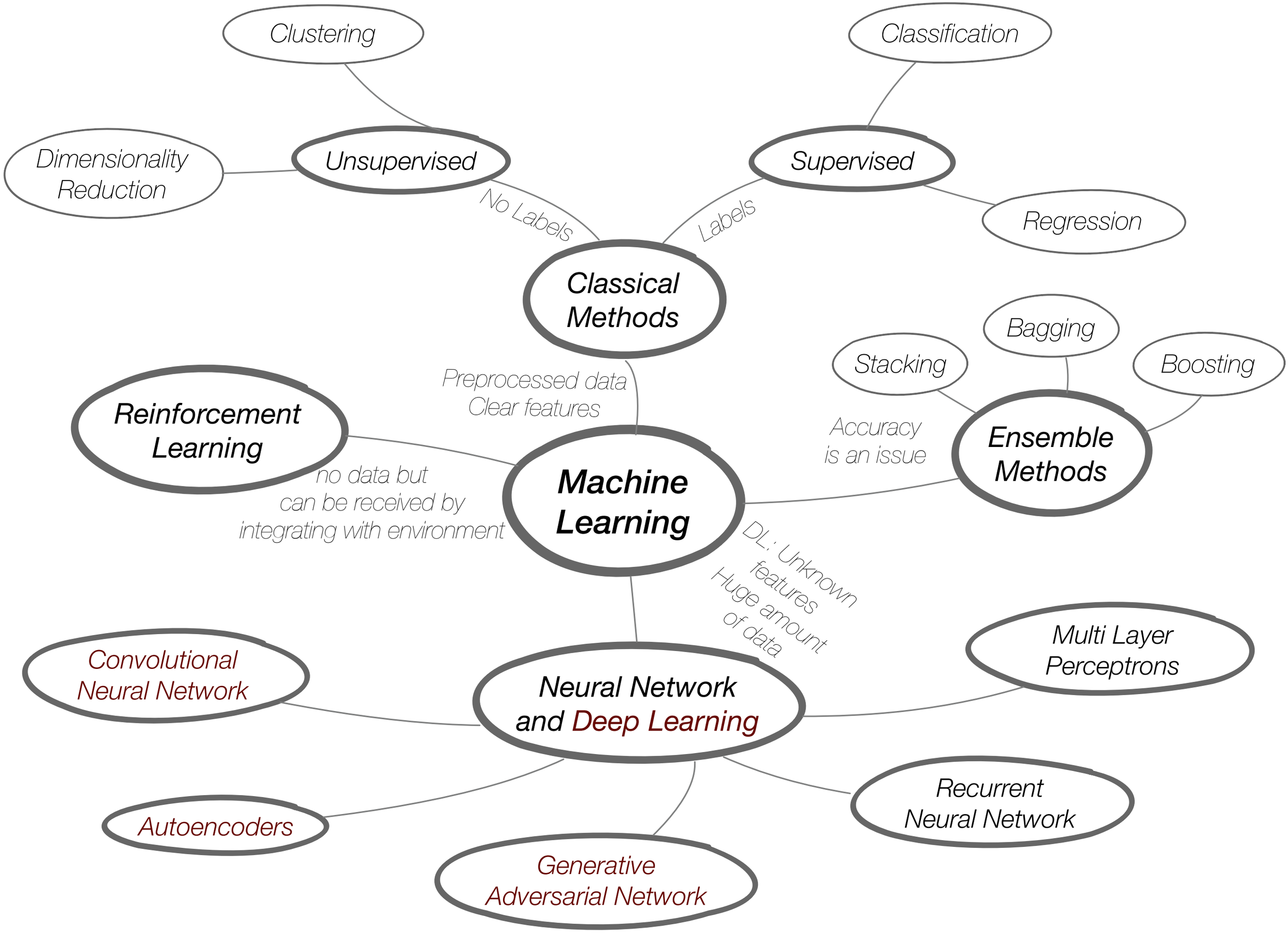
End of Video 1.2

Machine Learning Techniques

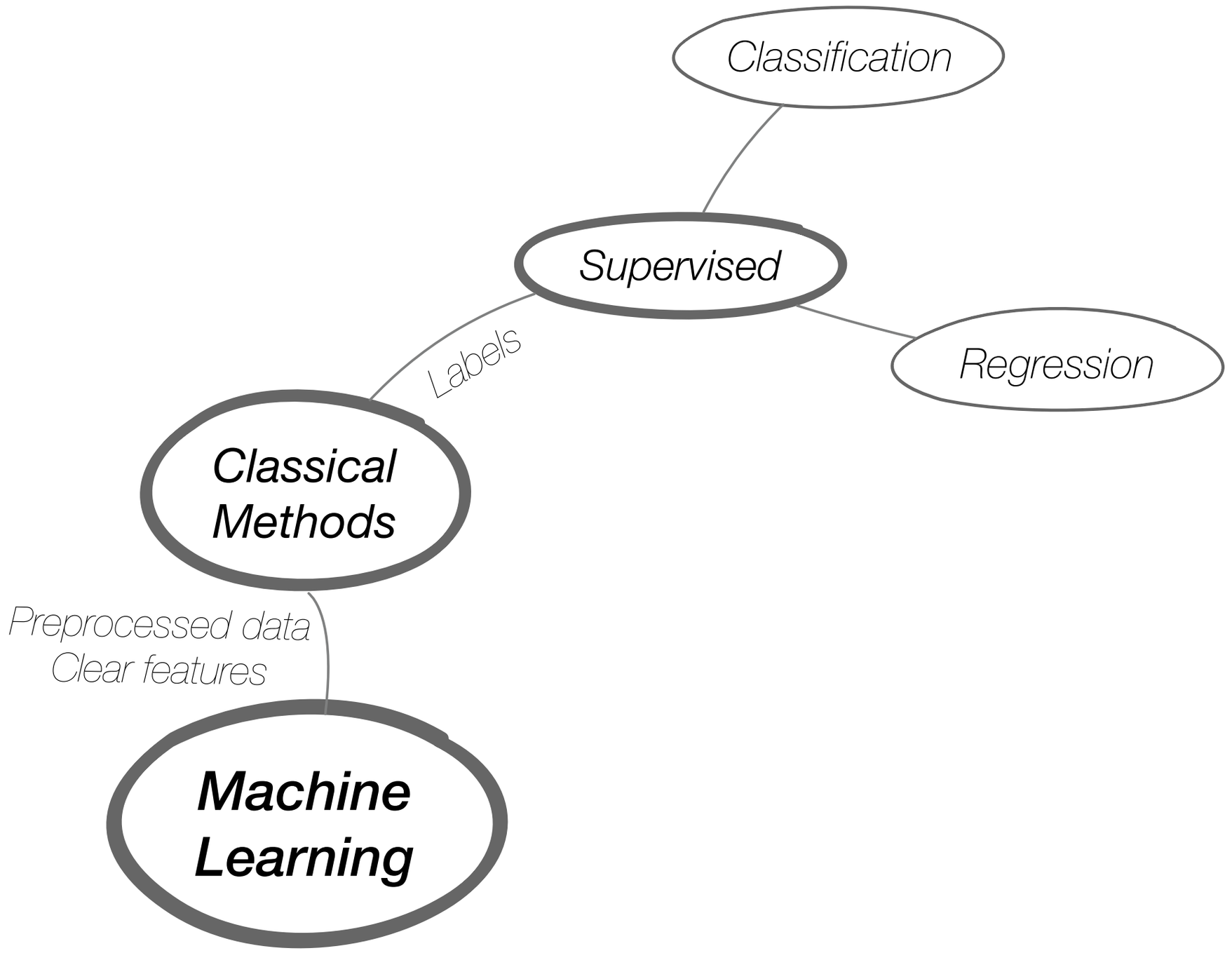
Lecture 1.3

- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning
- Ensemble Methods
- Neural Network
- Convolutional Neural Network


Clustering
Classification
Dimensionality Reduction
Unsupervised
Supervised
No Labels
Labels
Regression
Classical Methods
Bagging
Stacking
Boosting
Preprocessed data
Clear features
Reinforcement Learning
Accuracy is an issue
Ensemble Methods
no data but can be received by integrating with environment
Machine Learning
DL: Unknown features
Huge amount of data
Multi Layer Perceptrons
Convolutional Neural Network
Neural Network and Deep Learning
Recurrent Neural Network
Autoencoders
Generative Adversarial Network


Clustering
Classification
Dimensionality Reduction
Unsupervised
Supervised
Regression
No Labels
Labels
Classical Methods
Bagging
Stacking
Boosting
Preprocessed data
Clear features
Reinforcement Learning
Accuracy is an issue
Ensemble Methods
no data but
can be received by
integrating with environment
Machine Learning
DL : Unknown features
Huge amount of data
Multi Layer Perceptrons
Convolutional Neural Network
Neural Network and Deep Learning
Recurrent Neural Network
Autoencoders
Generative Adversarial Network

Supervised Learning

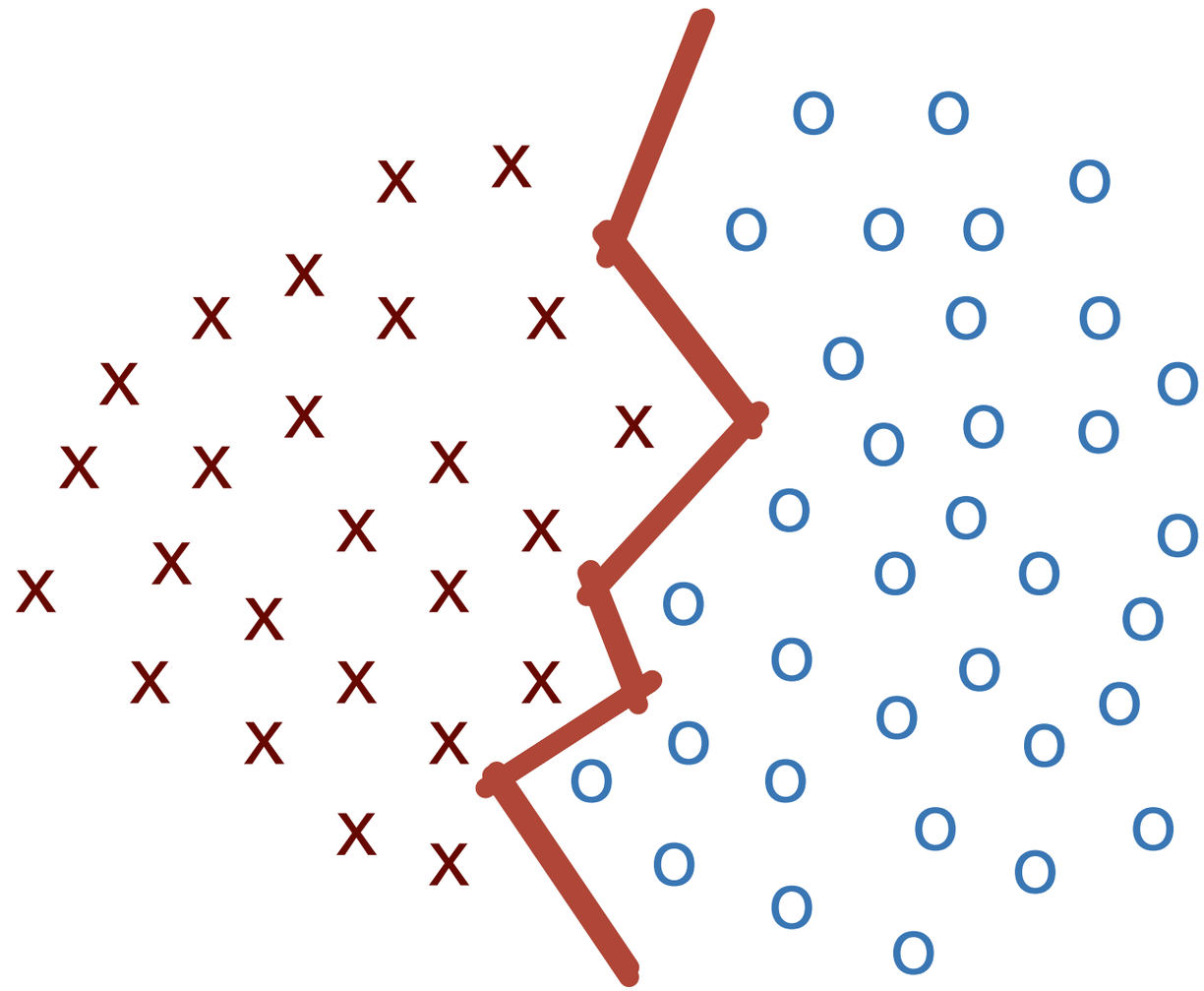


Supervised Learning Classification

"Splits objects based at one of the attributes known beforehand. Separate socks by based on color, documents based on language, music by genre"

Today used for:

- Spam filtering
- Language detection
- A search of similar documents
- Sentiment analysis
- Recognition of handwritten characters and numbers
- Fraud detection

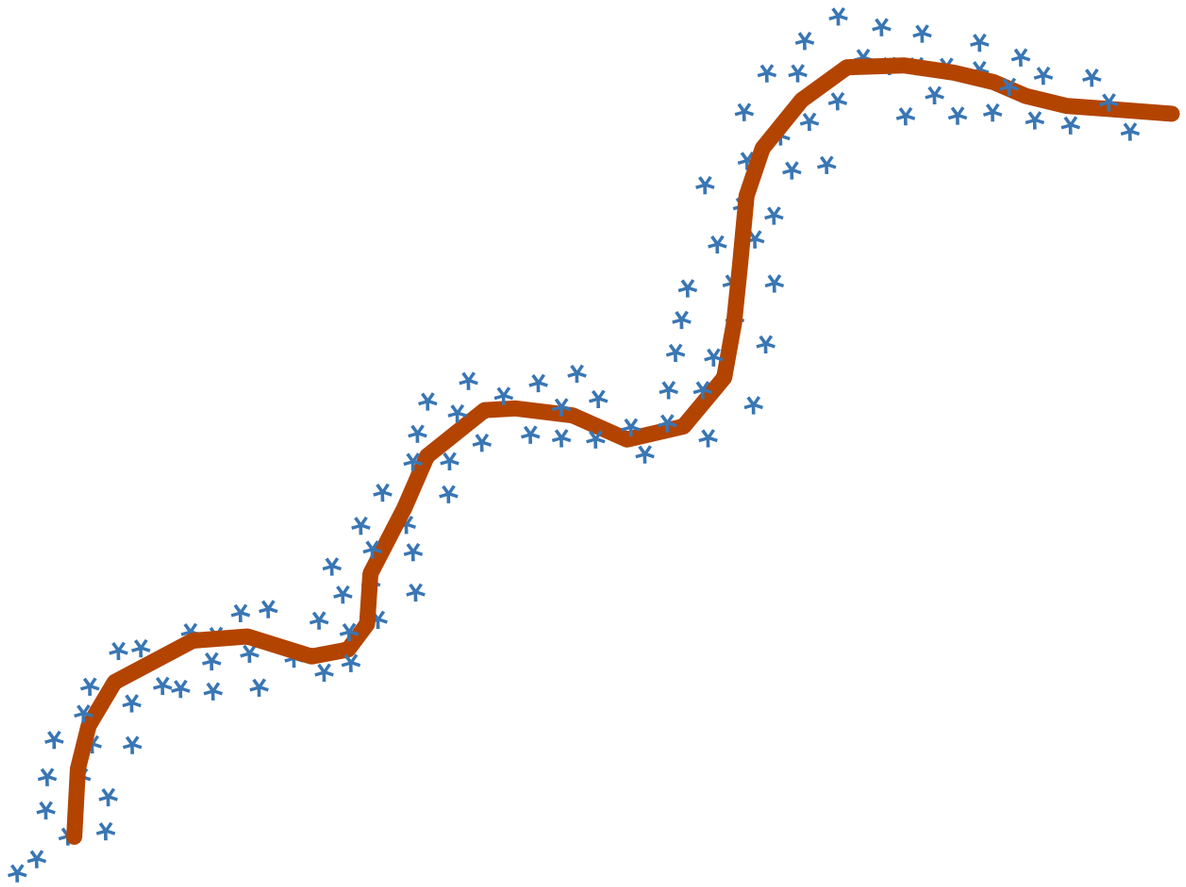


Popular algorithms: Naive Bayes, Decision Tree, Logistic Regression, K-Nearest Neighbours, Support Vector Machine

Slide source: https://vas3k.com/blog/machine_learning/?ref=hn&source=techstories.org#scroll70

Supervised Learning
Regression

"Draw a line through these dots. Divide the ties by length"



Today used for:

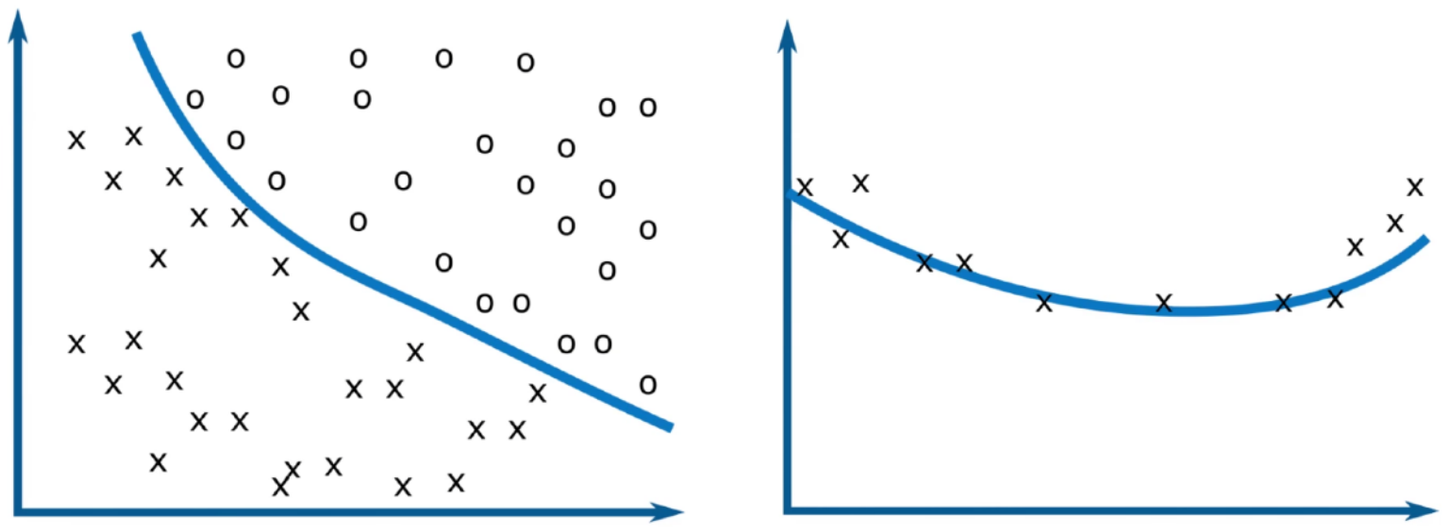
- Stock price forecasts
- Demand and sales volume analysis
- Medical diagnosis
- Any number-time correlations

Popular algorithms: Linear and Polynomial Regression

Slide source: https://vas3k.com/blog/machine_learning/?ref=hn&source=techstories.org#scroll70

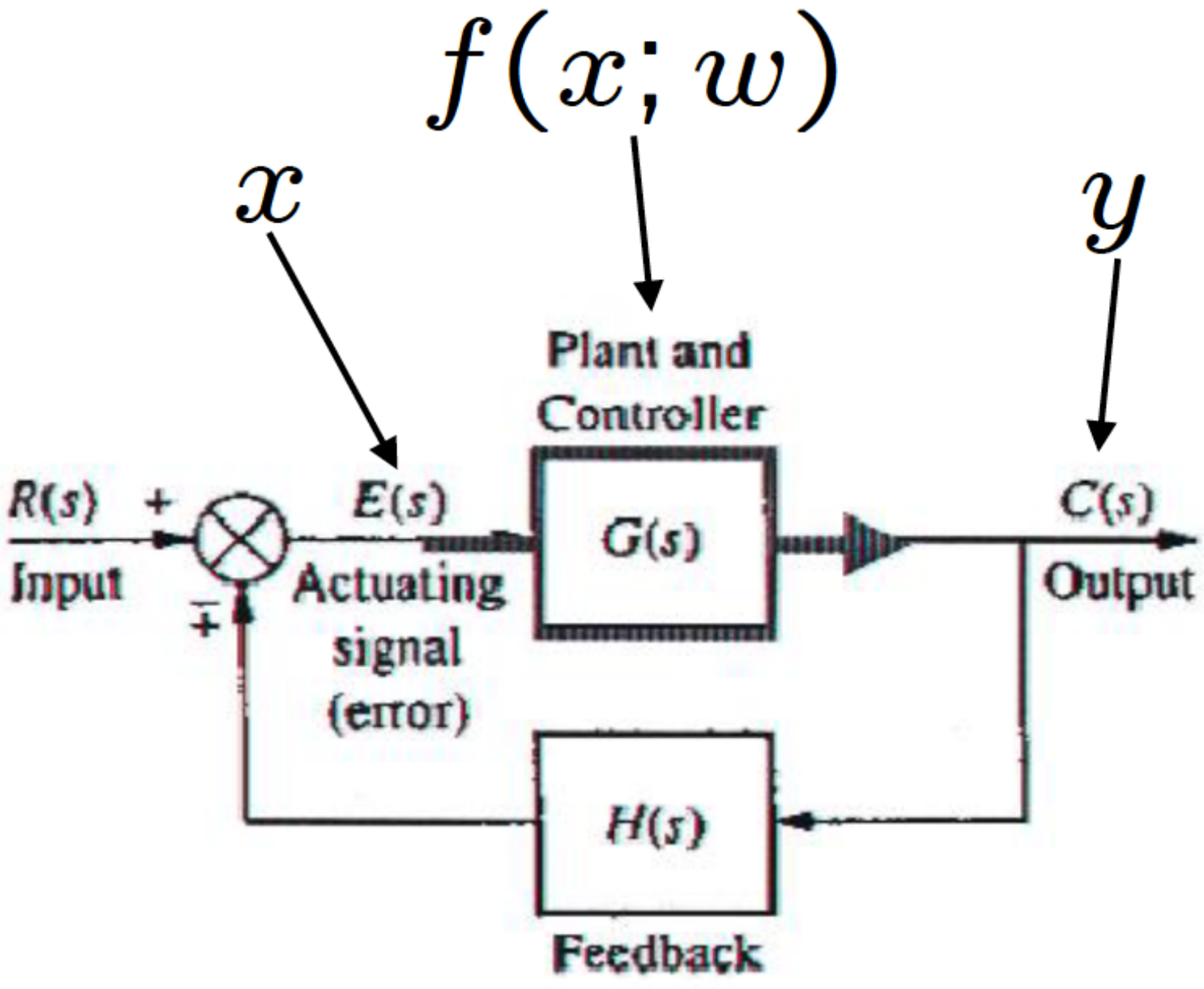
Classification vs. Regression

Property	Classification	Regression
Output type	Discrete (labels)	Continuous (number)
What do we search?	Decision boundary	Best fit line’
Evaluation	Accuracy	Sum of squared error



Pictures are taken from ‘Machine Learning Introduction: Regression and Classification | Intel Software’

Examples: Regression



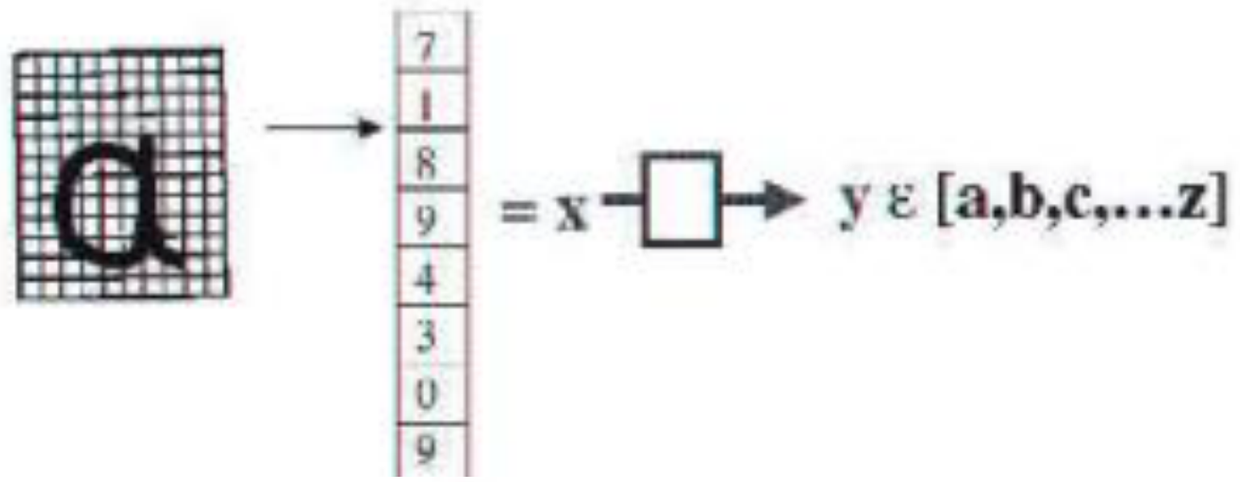
Slide source: B. Leibe Slides ML, 2015

Examples Classification

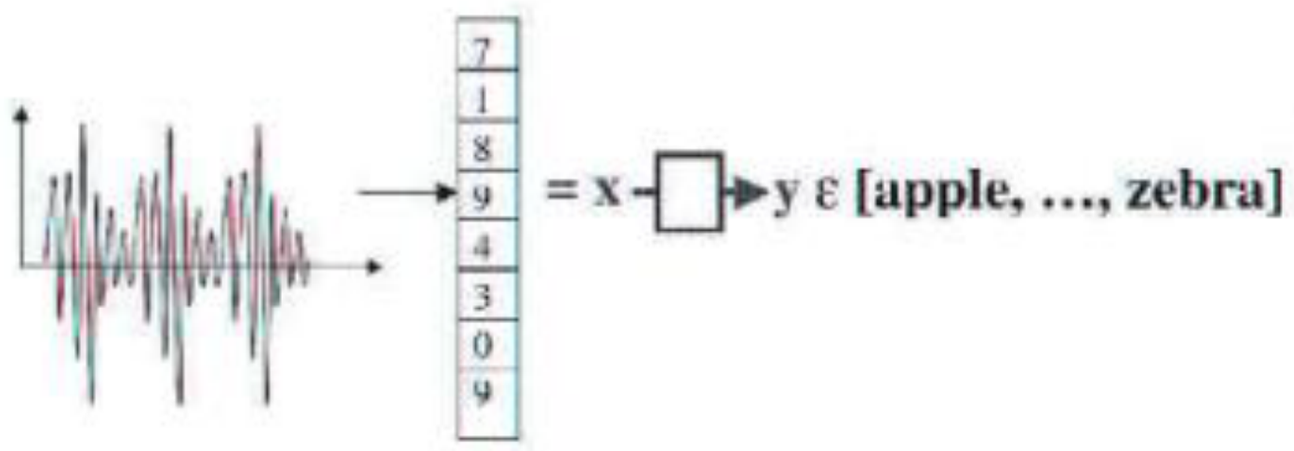
Email filtering

$$x \in [\text{a-z}]$$
$$y \in [\text{important}, \text{spam}]$$

Character recognition



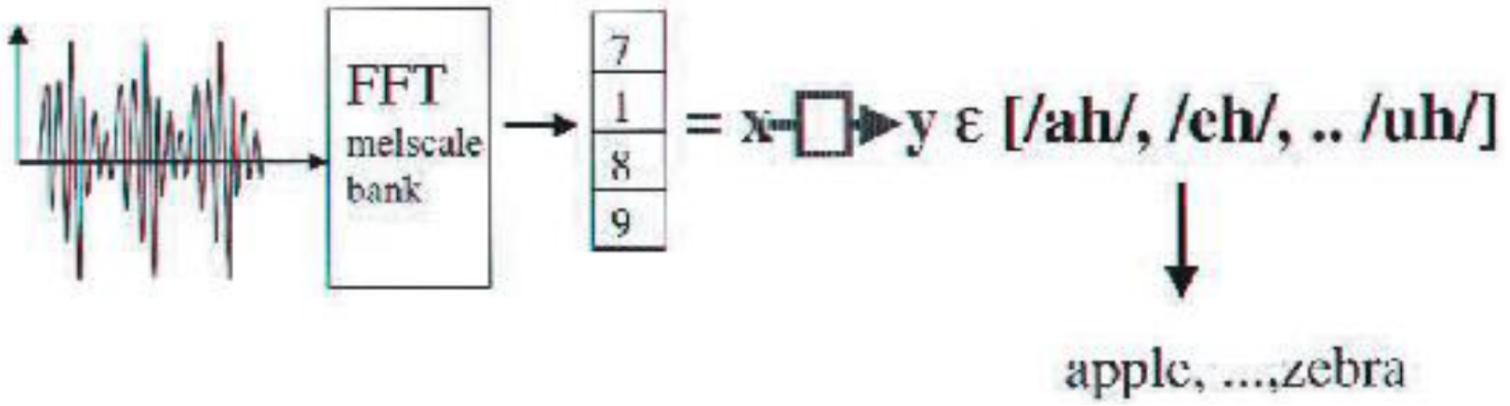
Speech recognition



Slide source: B. Leibe Slides ML, 2015

Machine Learning Core Problems

Input x



Features

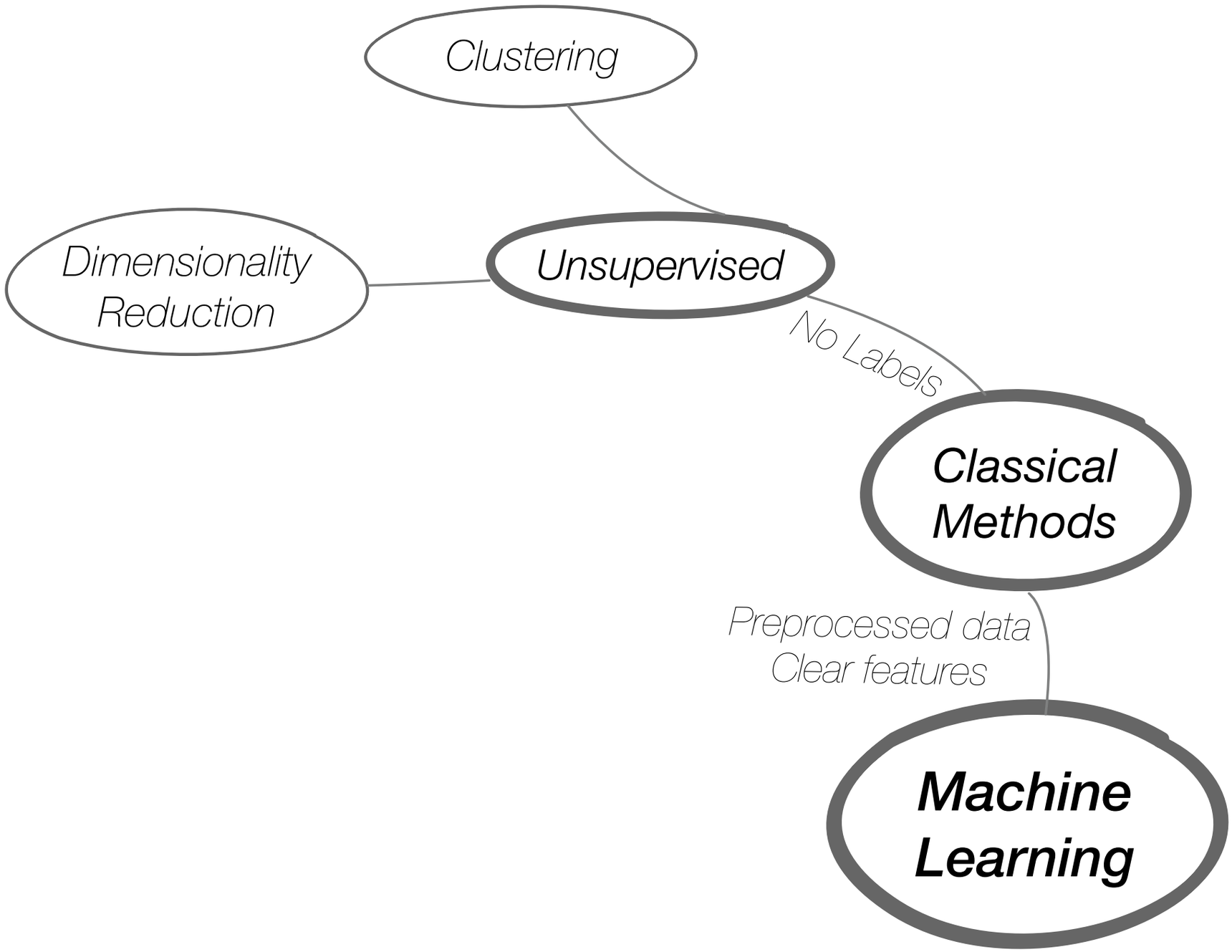
- Invariance to irrelevant input variations
- Selecting the ‘right’ features is crucial
- Encoding and use of ‘domain knowledge’
- Higher-dimensional features are more discriminative

Curse of dimensionality

- Complexity increases exponentially with number of dimensions

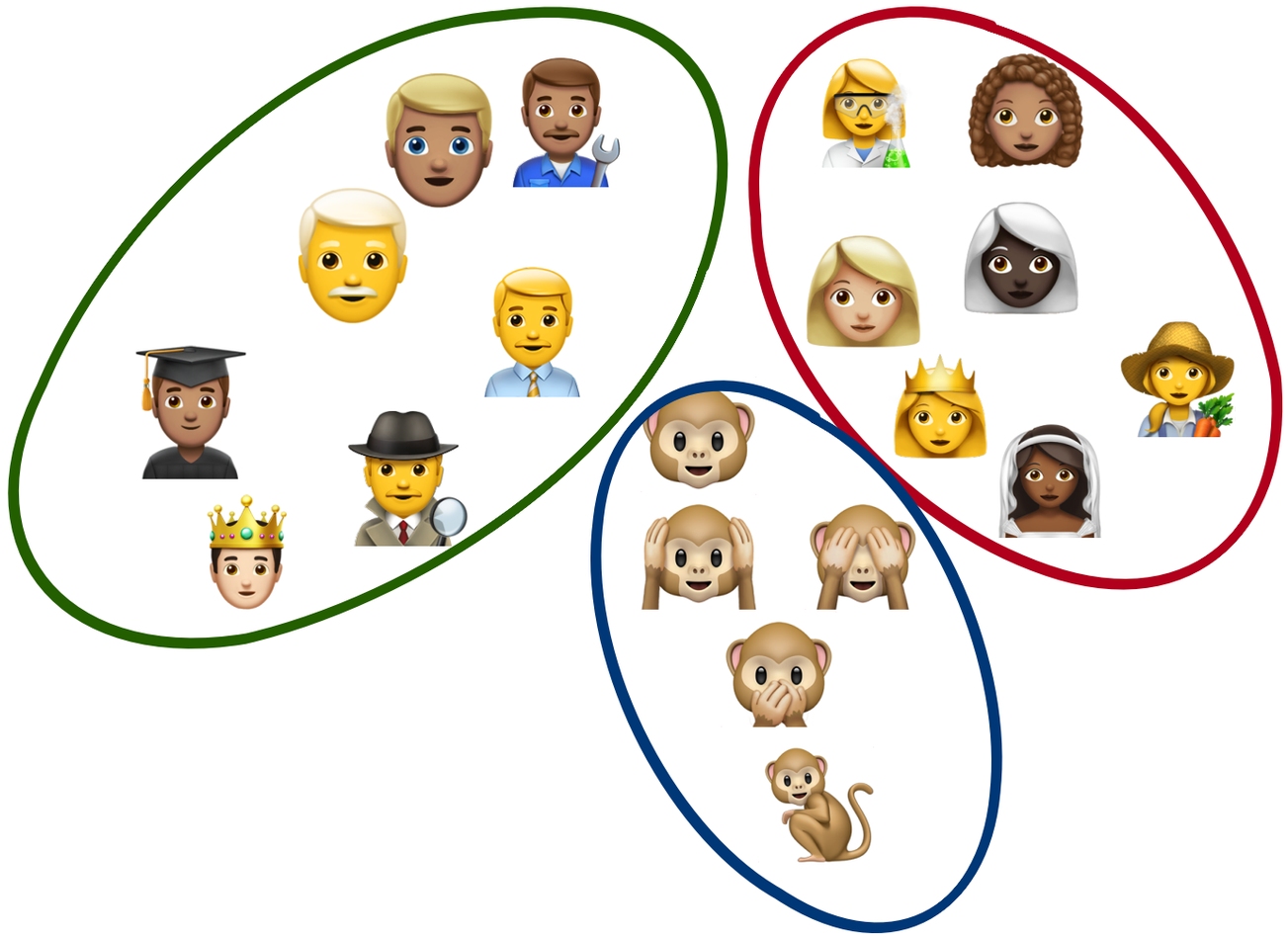
Slide source: B. Leibe Slides ML, 2015

Unsupervised Learning



Unsupervised Learning
Clustering

"Divides objects based on unknown features.
Machine chooses the best way"



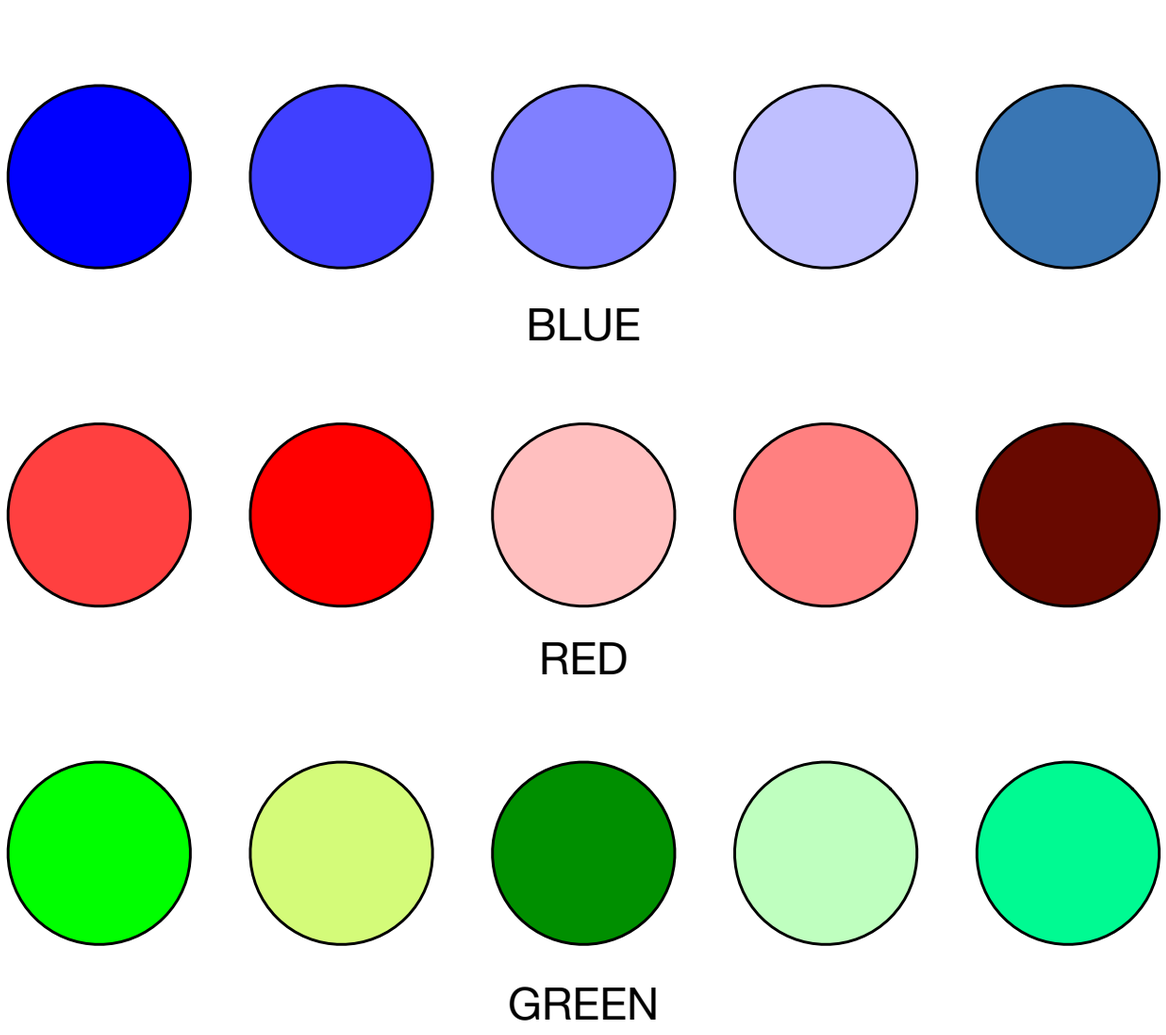
Nowadays used:

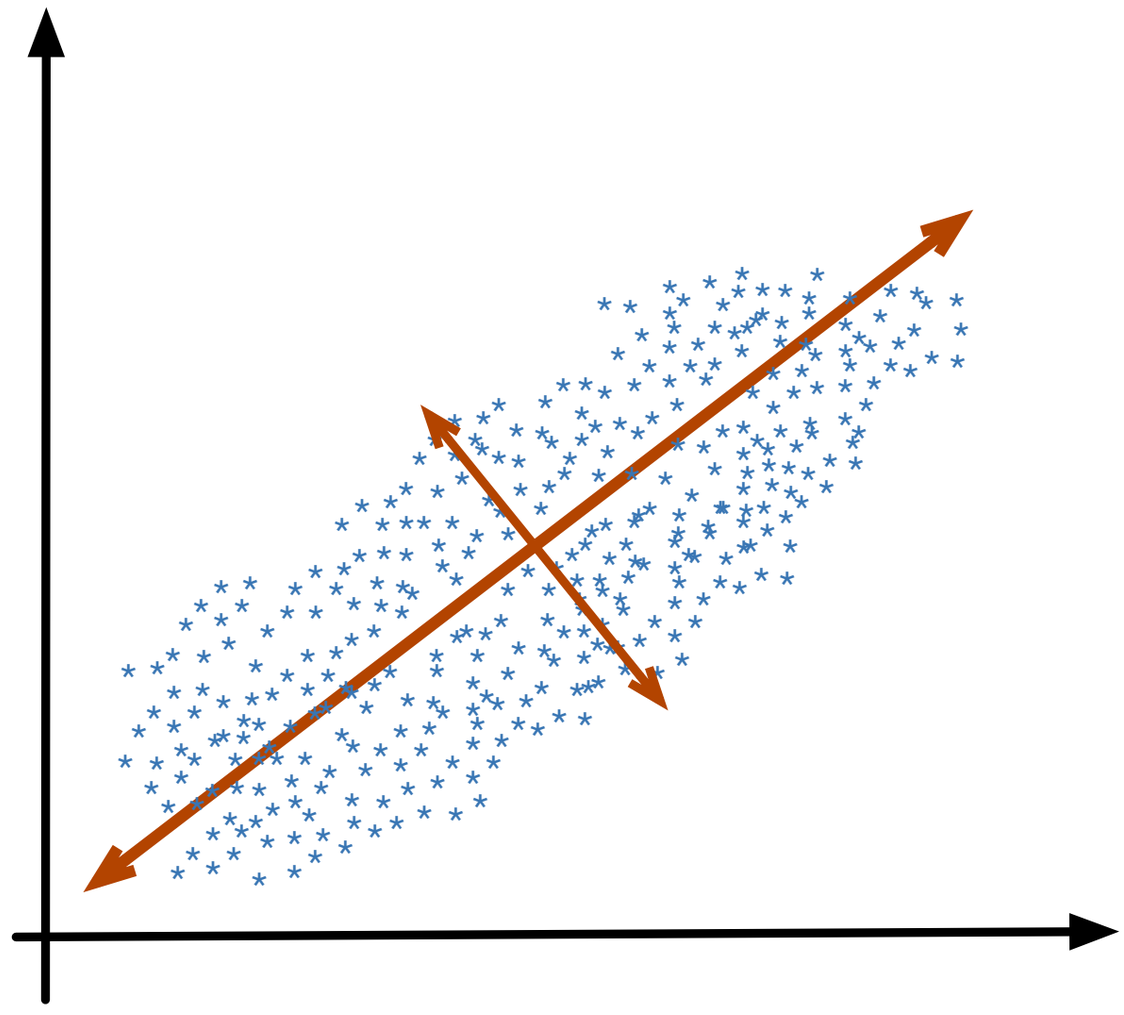
- For market segmentation (types of customers, loyalty)
- To merge close points on a map
- For image compression
- To analyze and label new data
- To detect abnormal behavior

Popular algorithms: K-means_clustering, Mean-Shift, DBSCAN

Slide source: https://vas3k.com/blog/machine_learning/?ref=hn&source=techstories.org#scroll70

Unsupervised Learning
Dimensionality Reduction





Assembles specific feature into more high level ones

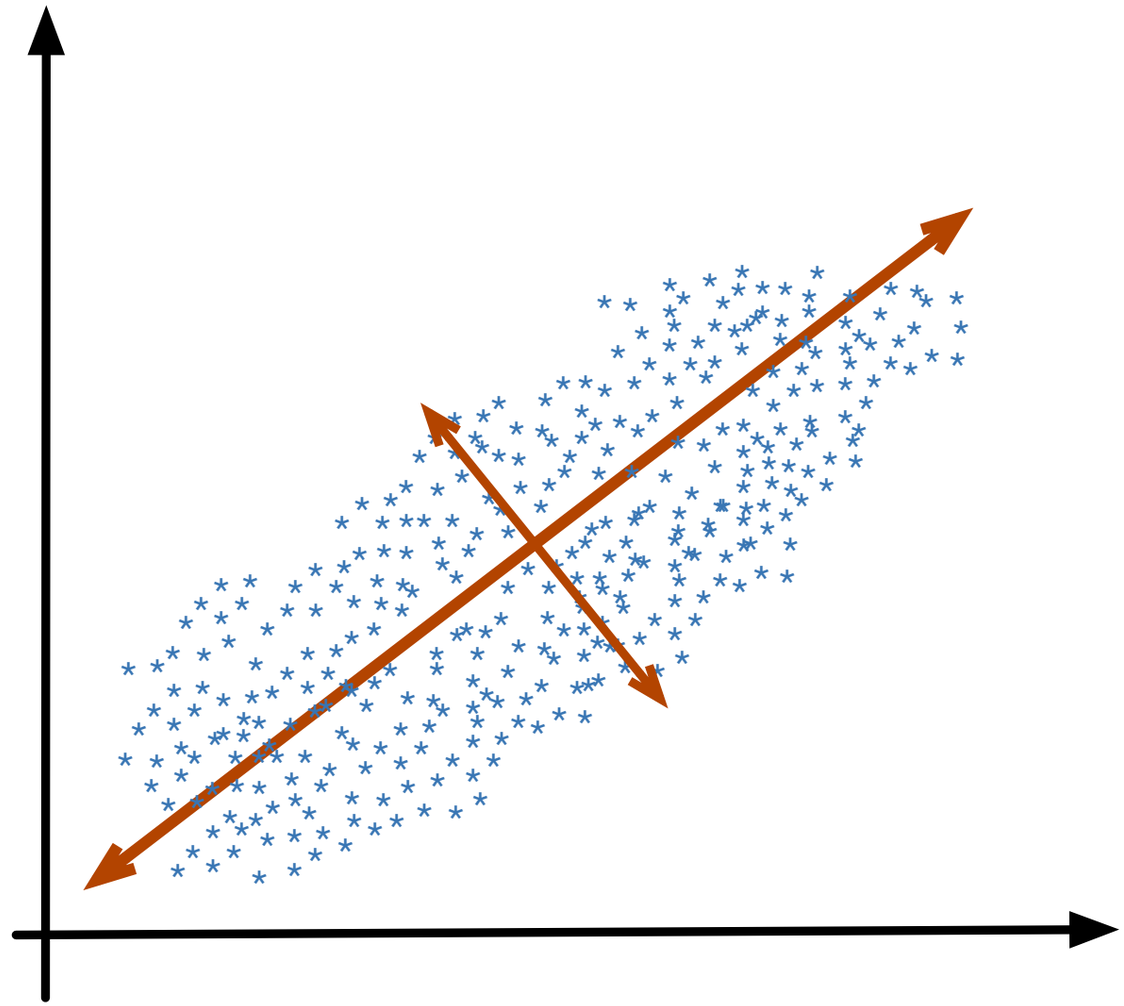
Slide source: https://vas3k.com/blog/machine_learning/?ref=hn&source=techstories.org#scroll70

Unsupervised Learning
Dimensionality Reduction

"Assembles specific features into more high-level ones"

Nowadays used for:

- Recommender systems
- Beautiful visualizations
- Topic modeling and similar document search
- Fake image analysis
- Risk management

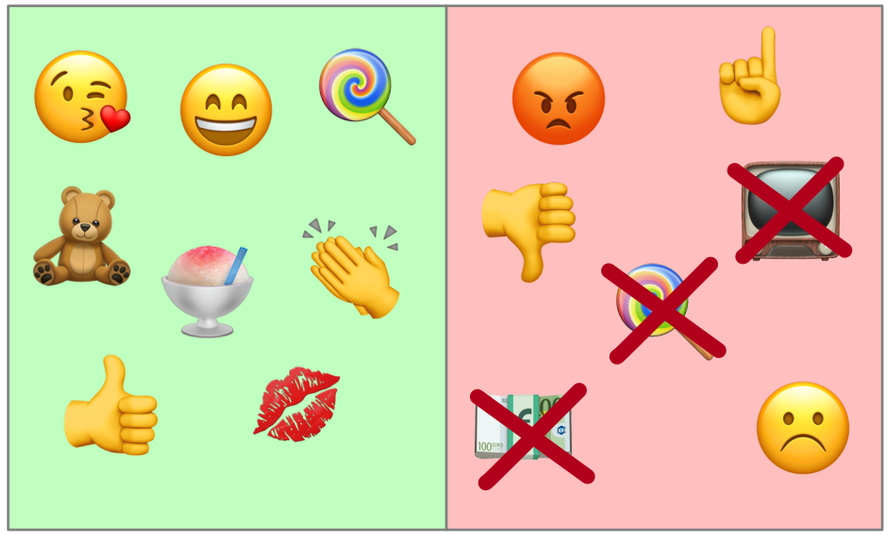


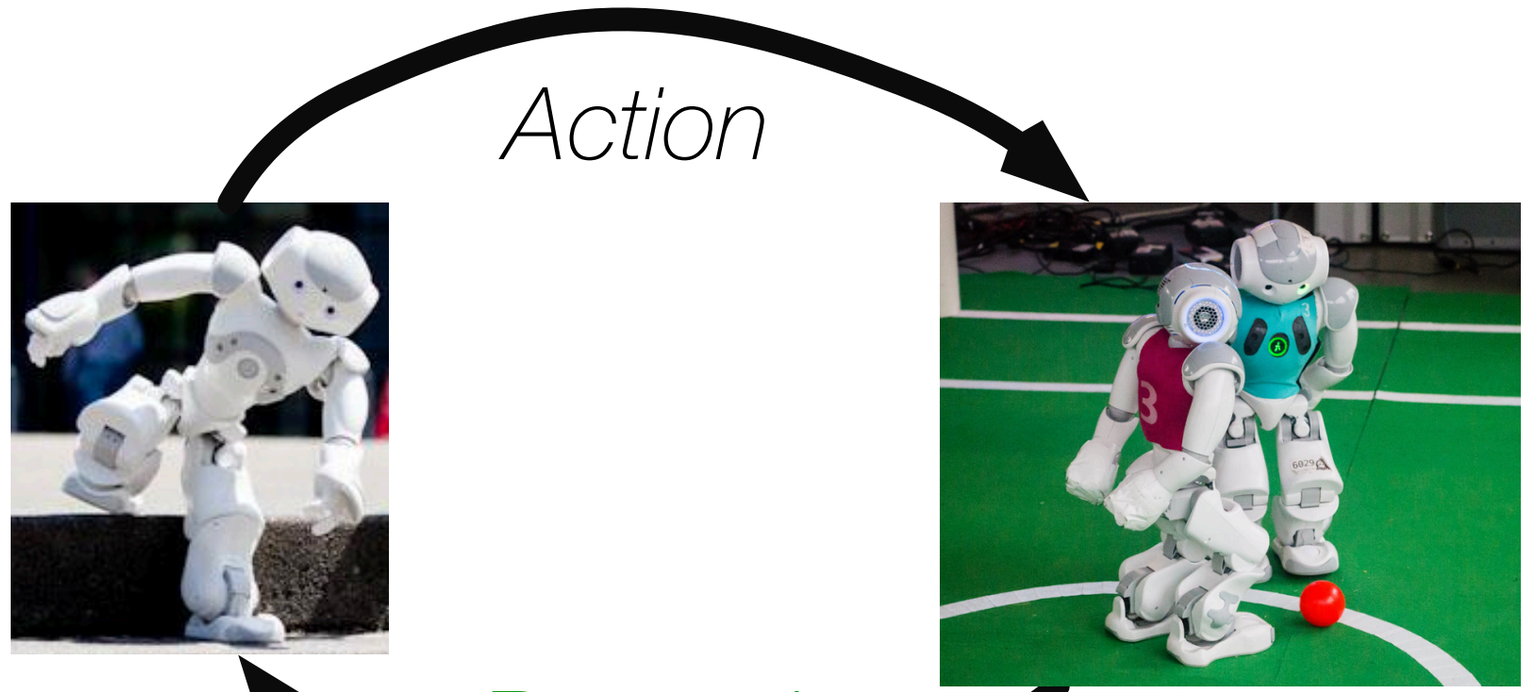
Popular algorithms: Principal Component Analysis (PCA), Singular Value Decomposition(SVD), Latent Dirichlet allocation (LDA), Latent Semantic Analysis (LSA, pLSA, GLSA), t-SNE (for visualization)

Slide source: https://vas3k.com/blog/machine_learning/?ref=hn&source=techstories.org#scroll70

Reinforcement Learning

Teach a robot to play soccer like we teach our children



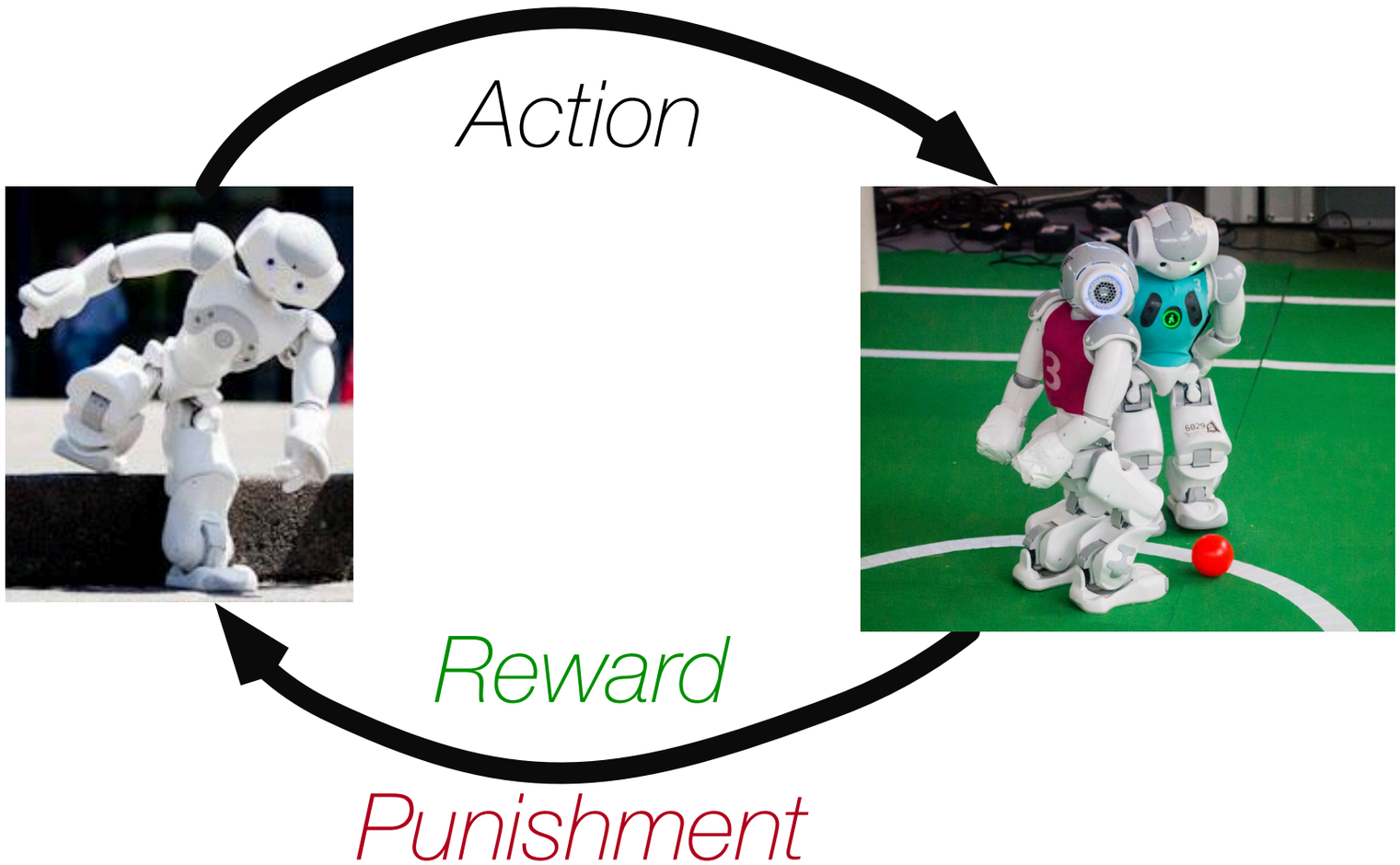


Reinforcement Learning

"Throw a robot into a maze and let it find an exit"

Nowadays used for:

- Self-driving cars
- Robot vacuums
- Games
- Automating trading
- Enterprise resource management



Popular algorithms: Q-Learning, SARSA, DQN, A3C, Genetic algorithm

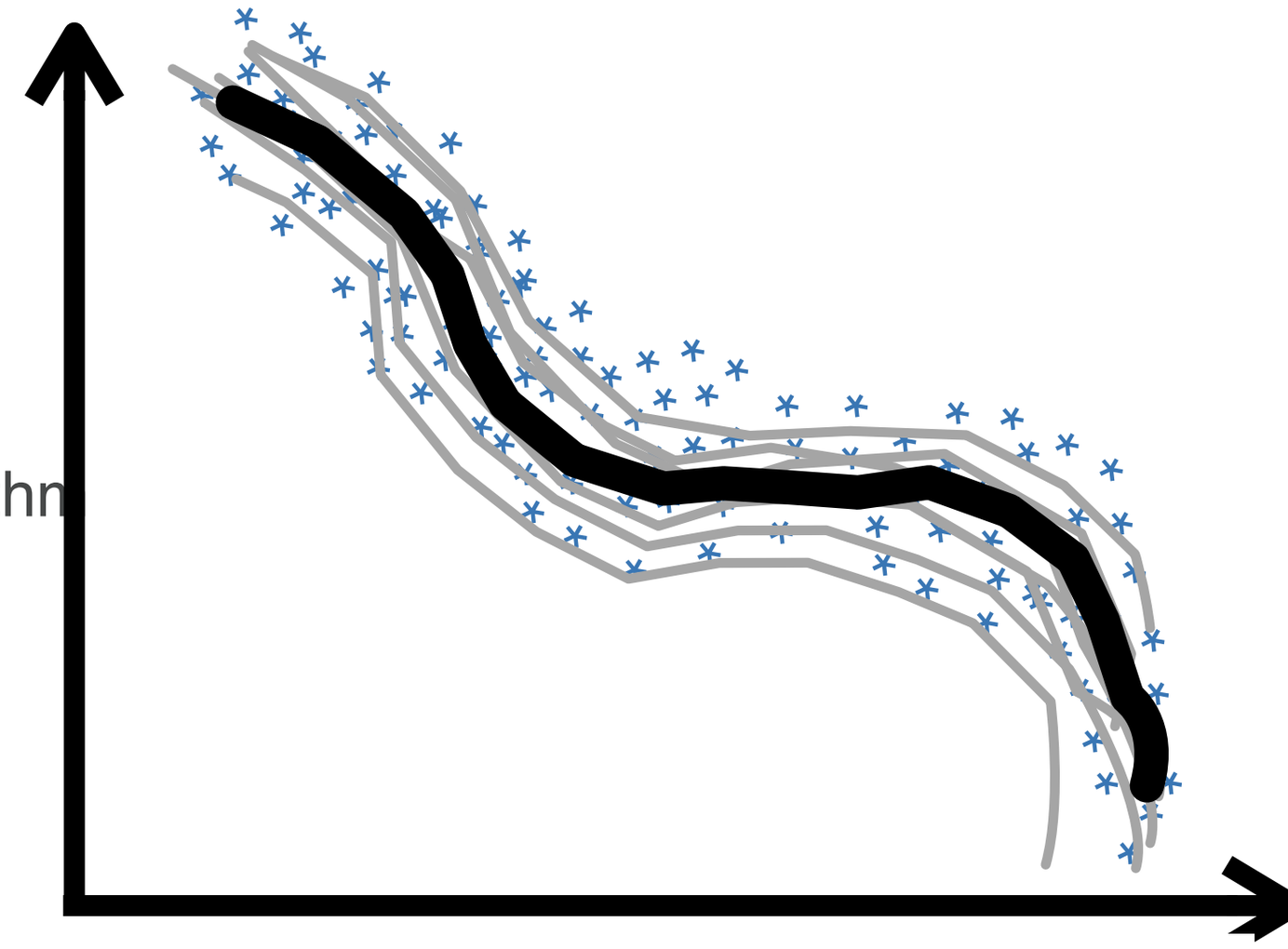
Slide source: https://vas3k.com/blog/machine_learning/?ref=hn&source=techstories.org#scroll70

Ensemble Methods

„Bunch of simple learning to correct errors of each other

Nowadays used for:

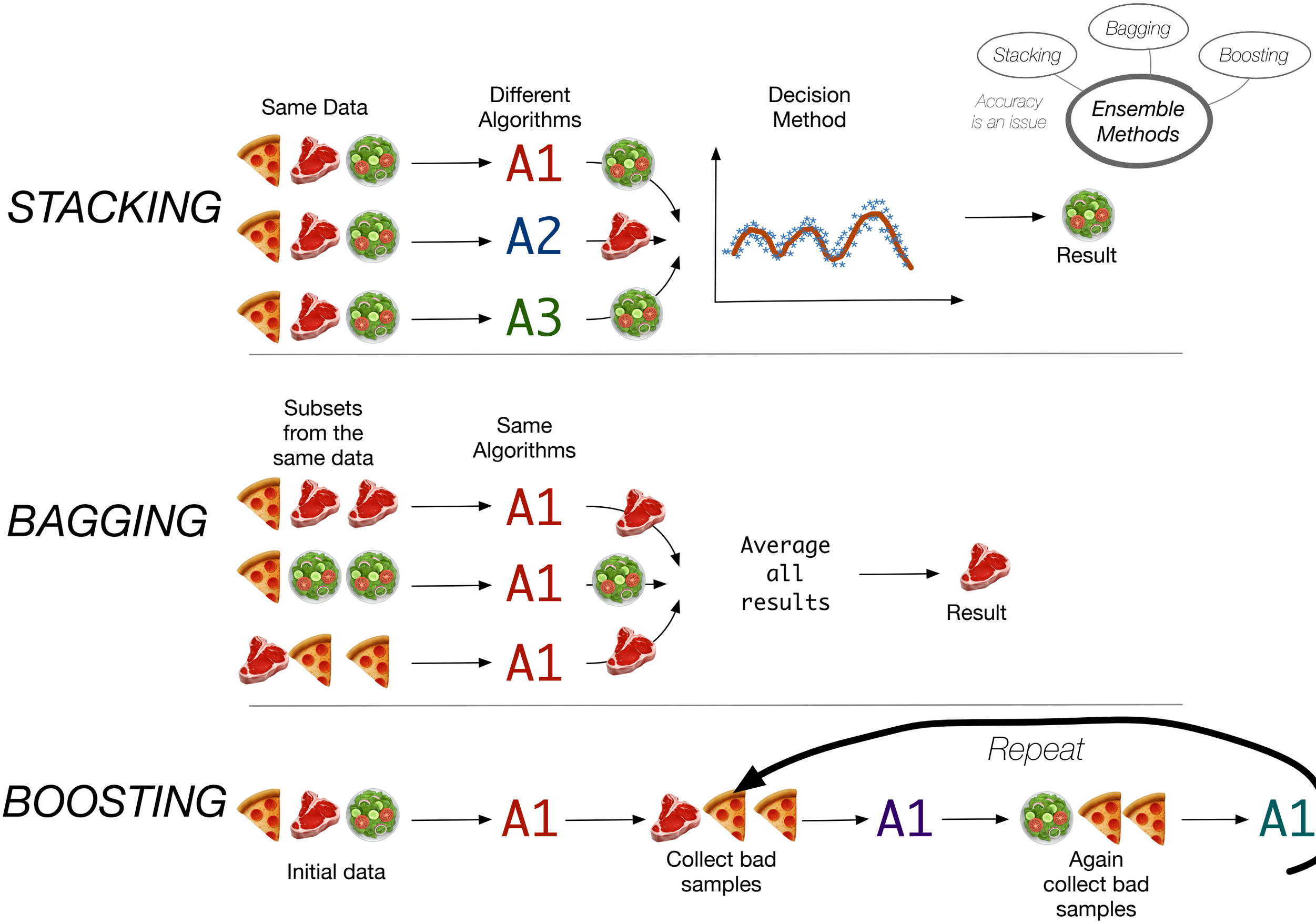
- Everything that fits classical algorithm approaches (but works better)
- Search systems
- Computer vision
- Object detection



Learning by averaging the outputs of several classifiers

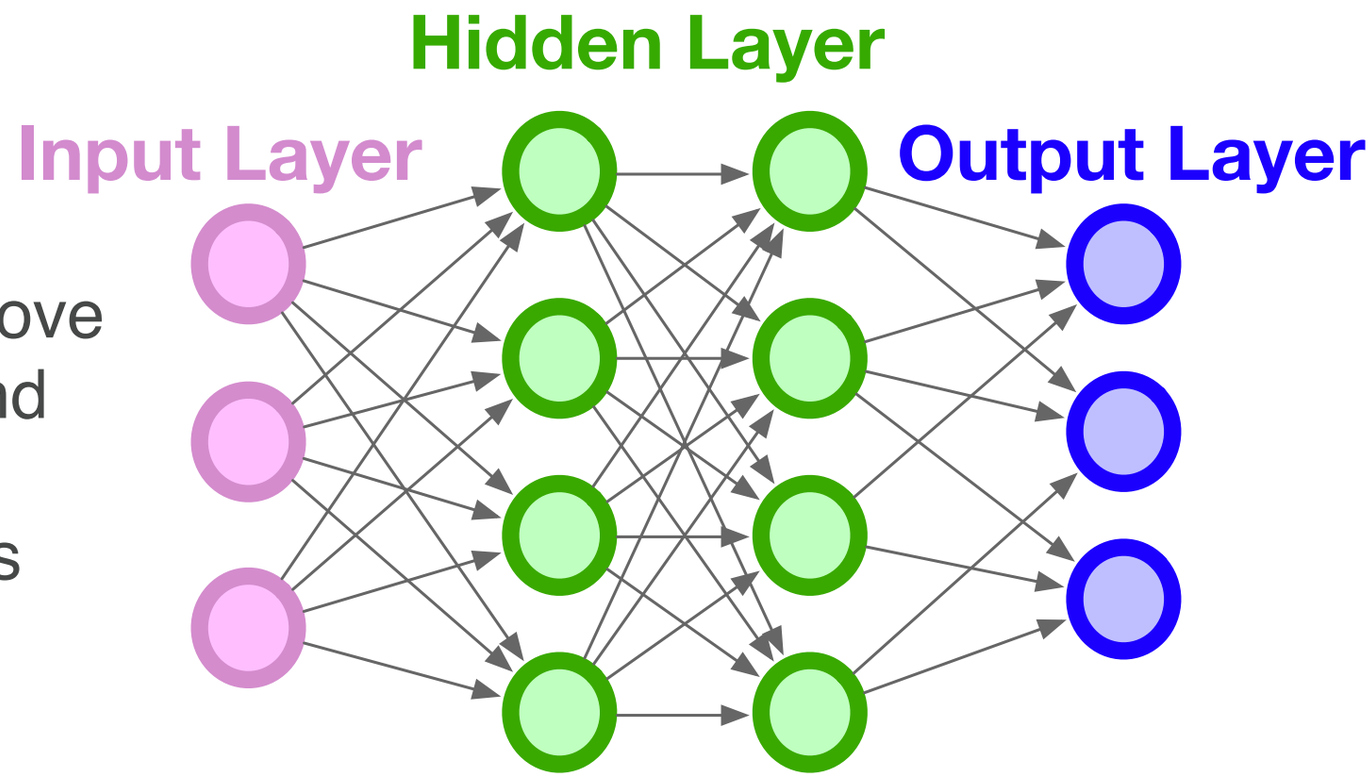
Popular algorithms: Random Forest, Gradient Boosting

Slide source: https://vas3k.com/blog/machine_learning/?ref=hn&source=techstories.org#scroll70


STACKING
Same Data
Different Algorithms
A1
A2
A3
Decision Method
Result
Bagging
Stacking
Boosting
Accuracy is an issue
Ensemble Methods
Subsets from the same data
Same Algorithms
BAGGING
A1
A1
A1
Average all results
Result
BOOSTING
Repeat
A1
A1
A1
Initial data
Collect bad samples
Again collect bad samples

Neural Network and Deep Learning

„We have a thousand-layer network, dozens of video cards, but still no idea where to use it. Let's generate cat pics!“



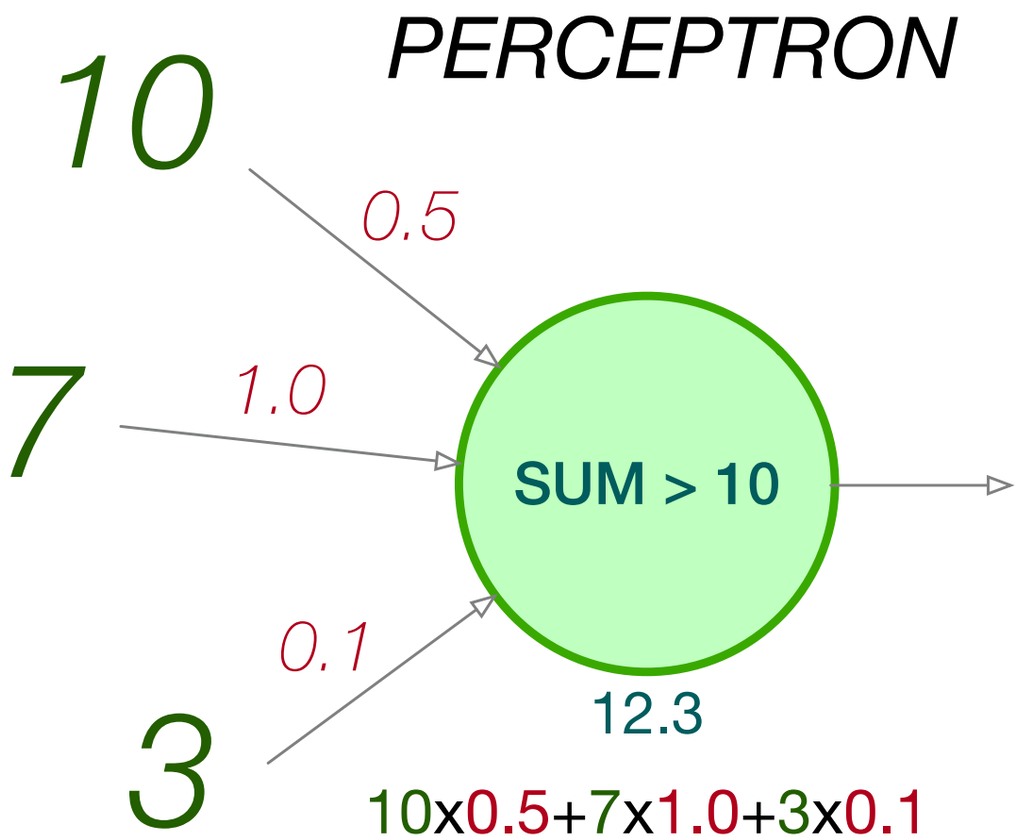
Nowadays used for:

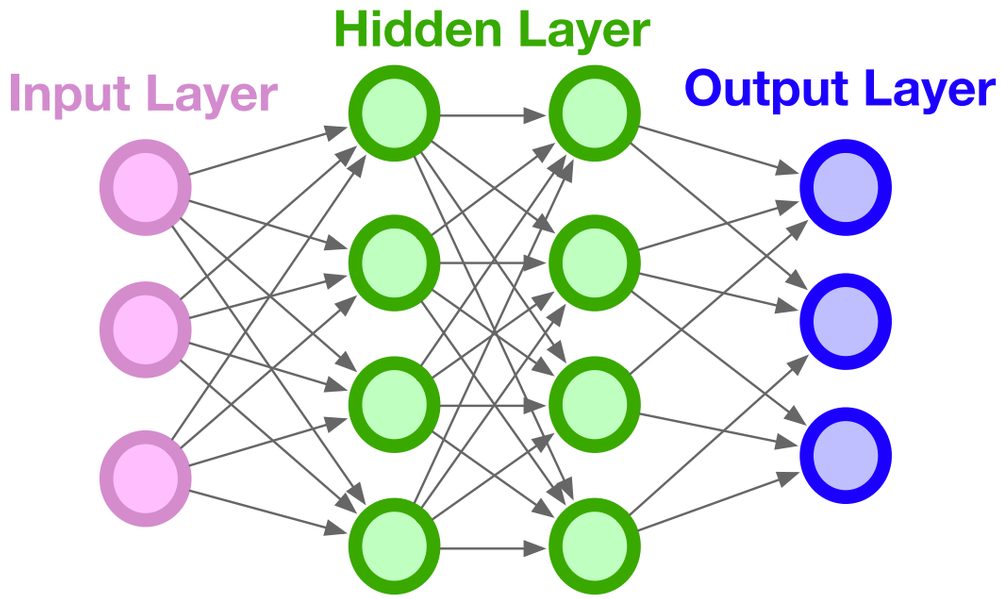
- Replacement of all algorithms above
- Object identification on photos and videos
- Speech recognition and synthesis
- Image processing, style transfer
- Machine translation

Popular algorithms: Perceptron, Convolutional Network (CNN), Recurrent Networks(RNN), Autoencoders

Slide source: https://vas3k.com/blog/machine_learning/?ref=hn&source=techstories.org#scroll70

Neural Network





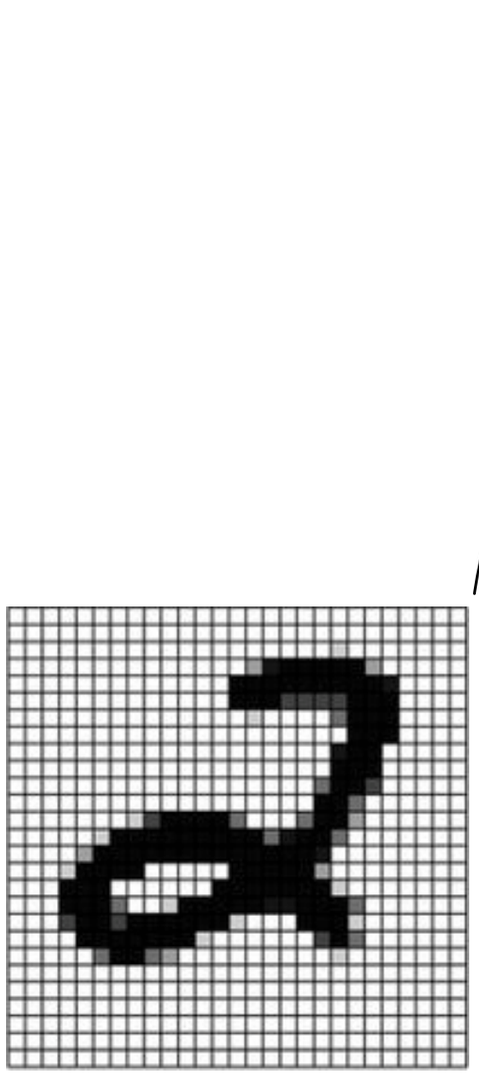


Image source: *Y. Jiang, J. Kang, X. Wang. RRAM-based parallel computing architecture using k-nearest neighbor classification for pattern recognition. Page 6.*

Inputs

pixel 1
pixel 2
pixel 3
pixel 4
pixel 5
pixel 6
pixel 779
pixel 780
pixel 781
pixel 782
pixel 783
pixel 784

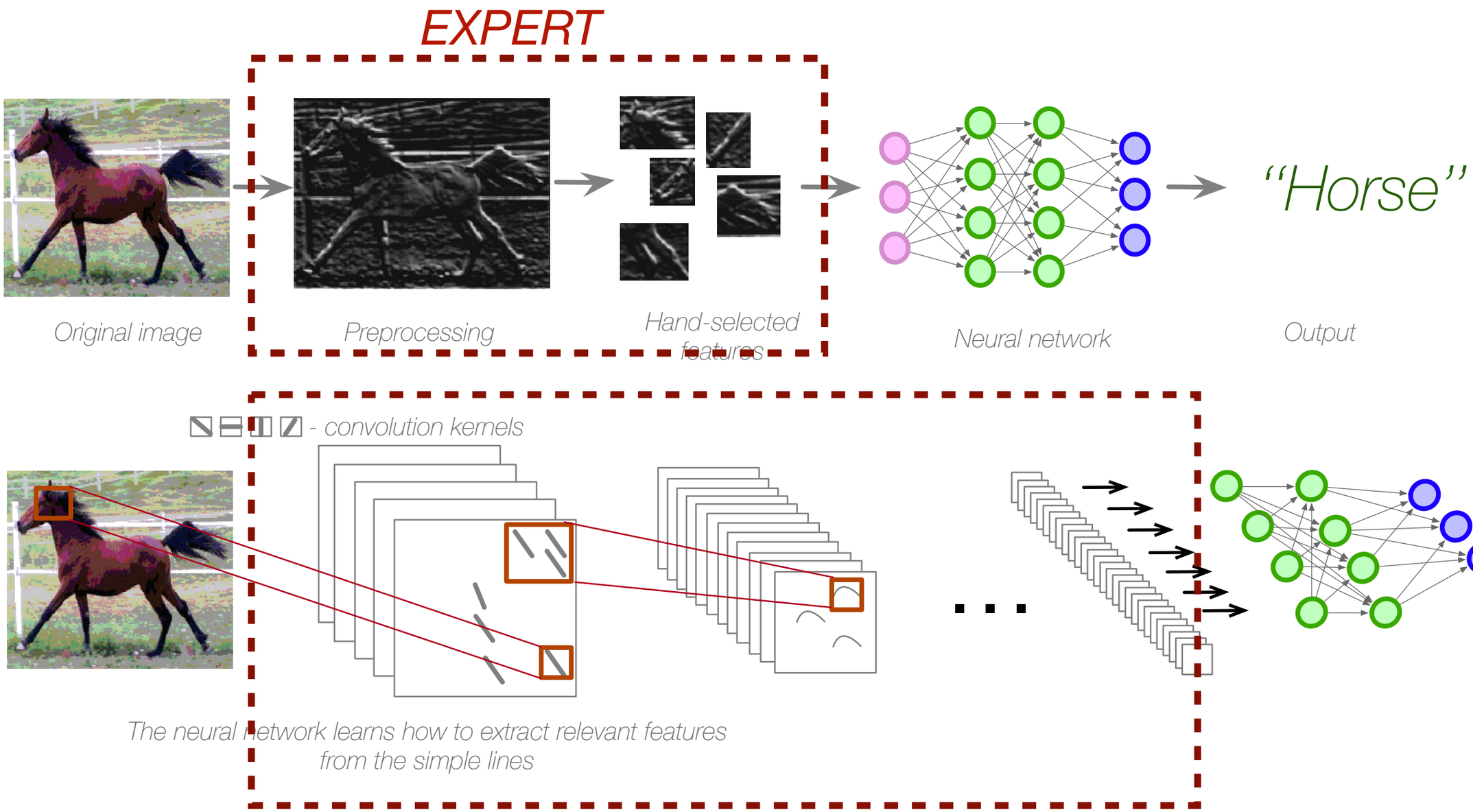
Hidden Layers

Outputs

0
1
2
3
4
5
6
7
8
9

2

Convolutional Neural Network



Course Outline

Basic Concepts

- Parametric Method,
- Bayesian Learning and Nonparametrics Methods

Classical Approaches

- Clustering and Mixture of Gaussians
- Linear Discriminants

Ensemble Methods

- Ensemble Methods and Boosting
- Randomized Trees, Forest

Reinforcement Learning

- Classical Reinforcement Learning

Neural Networks and Deep Learning

- Foundations
- Optimization

End of Video 1.3

Basic Concepts

Lecture 1.4

Fundamentals

- Recall of Probability Theory
- Probabilities
- Probability densities
- Expectations and covariances

Bayes Decision Theory

- Basic concepts
- Minimizing the misclassification rate
- Minimizing the expected loss

Probability Theory

"Probability theory is nothing but common sense reduced to calculation."

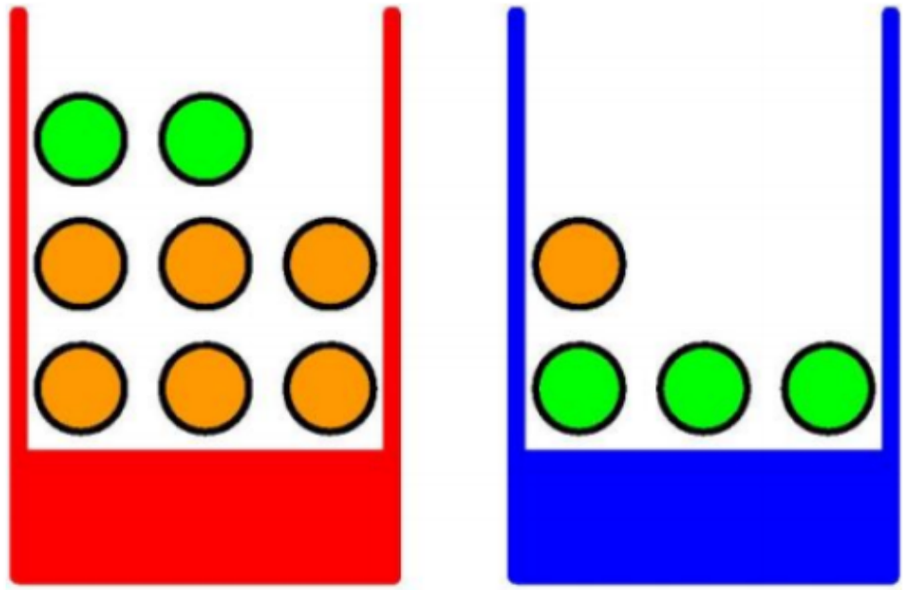
Pierre-Simon de Laplace, 1749-1827

Slide source: B. Leibe Slides ML, 2015

Probability Theory

Example: apples and oranges

- We have two boxes to pick from.
- Each box contains both types of fruit.
- What is the probability of picking an apple?



Formalization

- Let $B \in \{r,b\}$ be a random variable for the box we pick.
- Let $F \in \{a,o\}$ be a random variable for the type of fruit we get.
- Suppose we pick the red box 40% of the time. We write this as

$$p(B=r)=0.4 \qquad p(B=b)=0.6$$

- The probability of picking an apple given a choice for the box is

$$p(F=a|B=r)=0.25 \qquad p(F=a|B=b)=0.75$$

- What is the probability of picking an apple: $p(F=a)=?$

Slide source: B. Leibe Slides ML, 2015

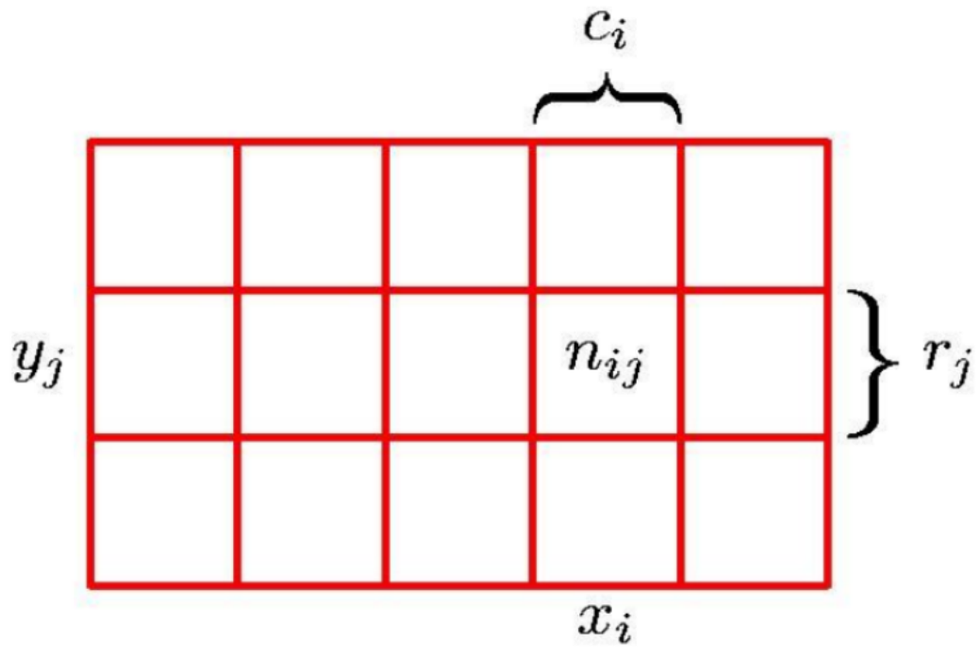
Probability Theory

More general case

Consider two random variables
$X \in \{x_i\}$ **and** $Y \in \{y_j\}$

Consider N trials and let

$$n_{ij} = \#\{X = x_i \wedge Y = y_j\}$$
$$c_i = \#\{X = x_i\}$$
$$r_j = \#\{Y = y_j\}$$



Then we can derive

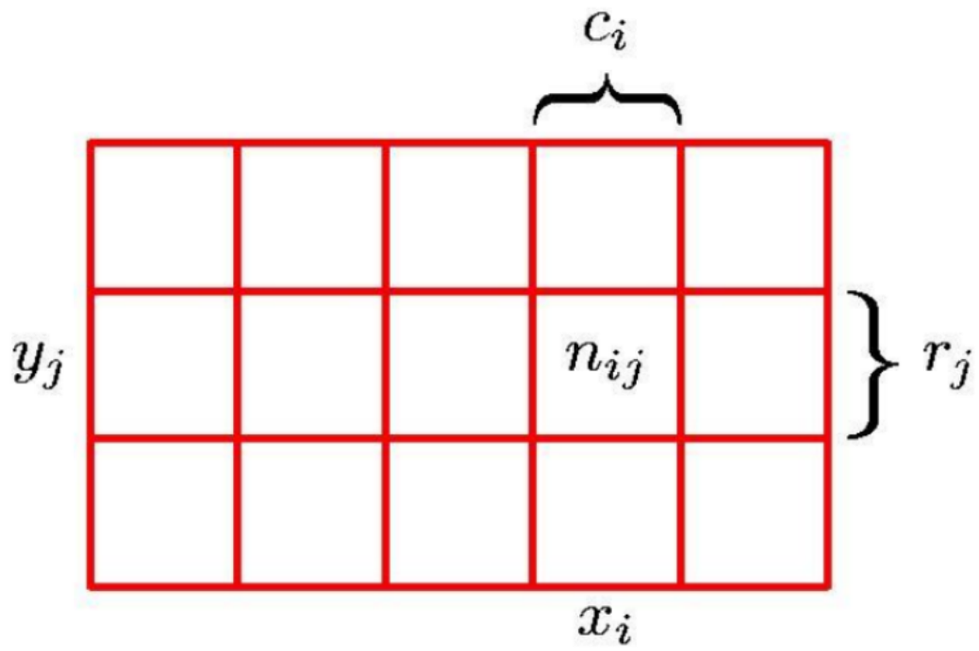
Joint probability
$$p(X = x_i, Y = y_j) = \frac{n_{ij}}{N}$$

Marginal probability
$$p(X = x_i) = \frac{c_i}{N}$$

Conditional probability
$$p(Y = y_j | X = x_i) = \frac{n_{ij}}{c_i}$$

Slide source: B. Leibe Slides ML, 2015

Probability Theory



Rules of probability

Sum Rule

$$p(X = x_i) = \frac{c_i}{N} = \frac{1}{N} \sum_{j=1}^{L} n_{ij} = \sum_{j=1}^{L} p(X = x_i, Y = y_j)$$

Product Rule

$$p(X = x_i, Y = y_j) = \frac{n_{ij}}{N} = \frac{n_{ij}}{c_i} \cdot \frac{c_i}{N}$$
$$= p(Y = y_j | X = x_i) p(X = x_i)$$

Slide source: B. Leibe Slides ML, 2015

The Rules of Probability

Thus we have

Sum Rule

$$p(X) = \sum_{Y} p(X, Y)$$

Product Rule

$$p(X, Y) = p(Y|X)p(X)$$

From these we can derive

Bayes' Theorem

$$p(Y|X) = \frac{p(X|Y)p(Y)}{p(X)}$$

where

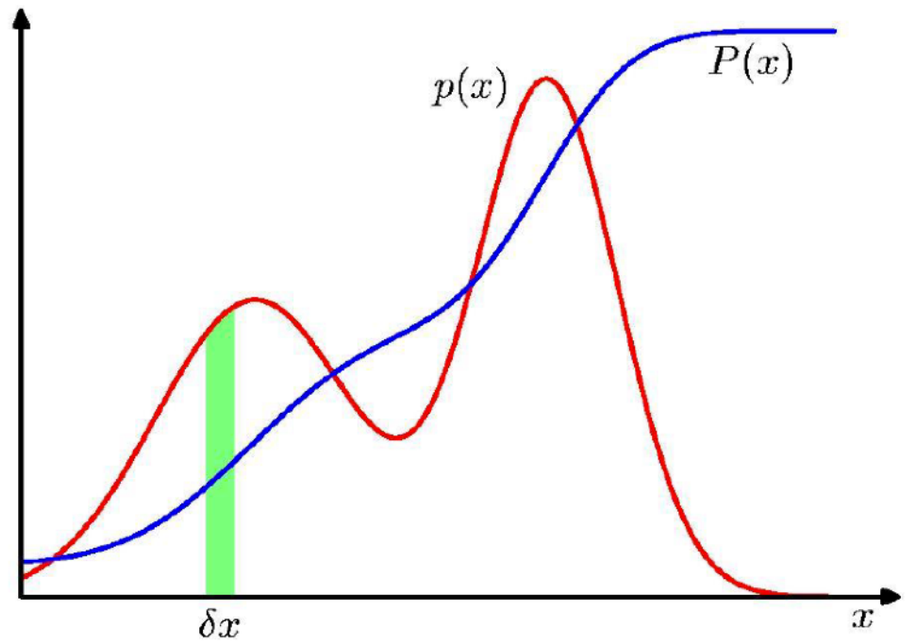
$$p(X) = \sum_{Y} p(X|Y)p(Y)$$

Slide source: B. Leibe Slides ML, 2015

Probability Densities

Probabilities over continuous variables are defined over their probability density function (pdf) *p(x)*

$$p(x \in (a,b)) = \int_a^b p(x)\,\mathrm{d}x$$



The probability that x lies in the interval *(-inf,z)* is given by the cumulative distribution function

$$P(z) = \int_{-\infty}^{z} p(x)\,\mathrm{d}x$$

Slide source: B. Leibe Slides ML, 2015

Expectations

The average value of some function *f(x)* under a probability distribution *p(x)* is called its **expectation**

$$\mathbb{E}[f] = \sum_x p(x) f(x) \qquad \mathbb{E}[f] = \int p(x) f(x)\, \mathrm{d}x$$

discrete case — continuous case

If we have a finite number N of samples drawn from a pdf, then the expectation can be approximated by

$$\mathbb{E}[f] \simeq \frac{1}{N} \sum_{n=1}^{N} f(x_n)$$

We can also consider a **conditional expectation**

$$\mathbb{E}_x[f|y] = \sum_x p(x|y) f(x)$$

Slide source: B. Leibe Slides ML, 2015

Variances and Covariances

The **variance** provides a measure how much variability there is in *f(x)* around its mean value E[*f(x)*].

$$\mathrm{var}[f] = \mathbb{E}\left[(f(x) - \mathbb{E}[f(x)])^2\right] = \mathbb{E}[f(x)^2] - \mathbb{E}[f(x)]^2$$

For two random variables x and y, the **covariance** is defines by

$$\begin{aligned}\mathrm{cov}[x, y] &= \mathbb{E}_{x,y}\left[\{x - \mathbb{E}[x]\}\{y - \mathbb{E}[y]\}\right] \\ &= \mathbb{E}_{x,y}[xy] - \mathbb{E}[x]\mathbb{E}[y]\end{aligned}$$

If x and y are vectors then the result is **covariance matrix**

$$\begin{aligned}\mathrm{cov}[\mathbf{x}, \mathbf{y}] &= \mathbb{E}_{\mathbf{x},\mathbf{y}}\left[\{\mathbf{x} - \mathbb{E}[\mathbf{x}]\}\{\mathbf{y}^{\mathrm{T}} - \mathbb{E}[\mathbf{y}^{\mathrm{T}}]\}\right] \\ &= \mathbb{E}_{\mathbf{x},\mathbf{y}}[\mathbf{x}\mathbf{y}^{\mathrm{T}}] - \mathbb{E}[\mathbf{x}]\mathbb{E}[\mathbf{y}^{\mathrm{T}}]\end{aligned}$$

Slide source: B. Leibe Slides ML, 2015

End of Video 1.4

Bayes Decision Theory

Lecture 1.5

- Basic concepts
- Minimizing the misclassification rate
- Minimizing the expected loss

Bayes Decision Theory

Thomas Bayes, 1701-1761

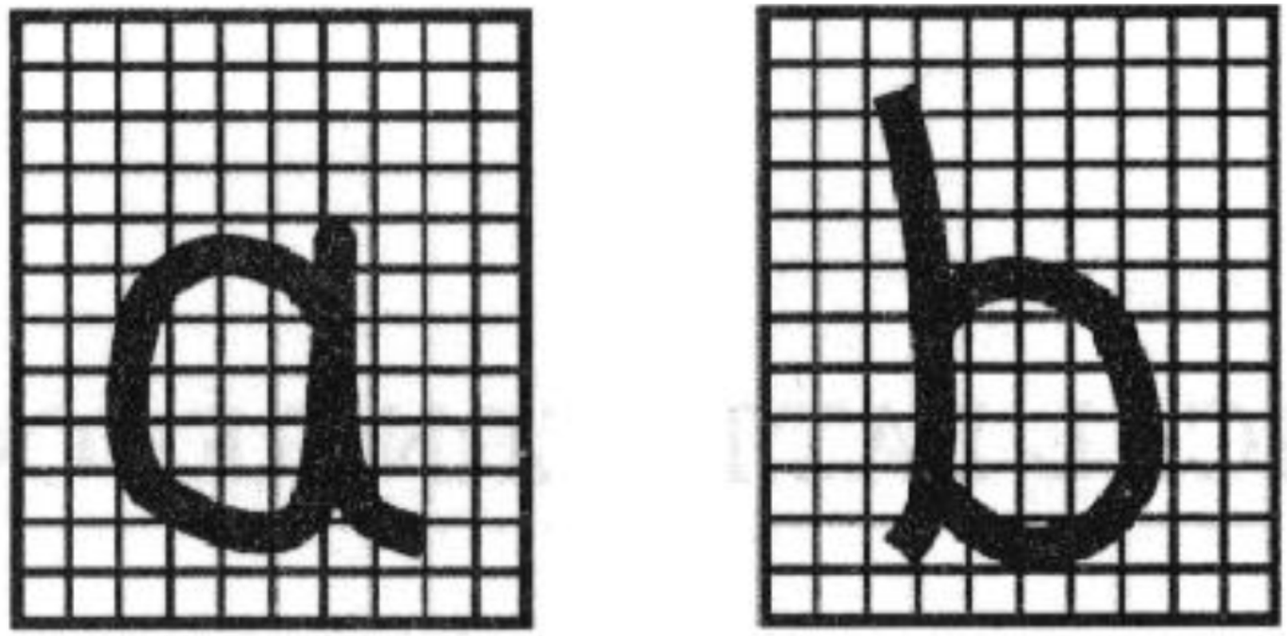
"The theory of inverse probability is founded upon an error, and must be wholly rejected."

R.A. Fisher, 1925

Slide source: B. Leibe Slides ML, 2015

Bayes Decision Theory

Example: handwritten character recognition



Goal:

Classify a new letter such that the probability of misclassification is minimised

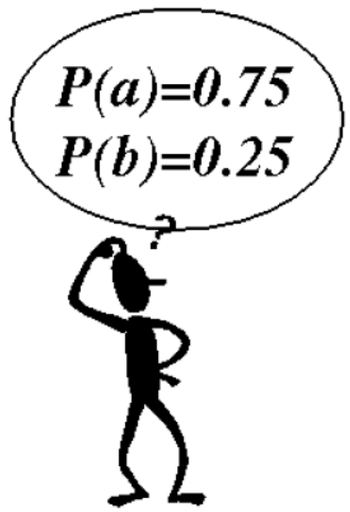
Slide source: B. Leibe Slides ML, 2015

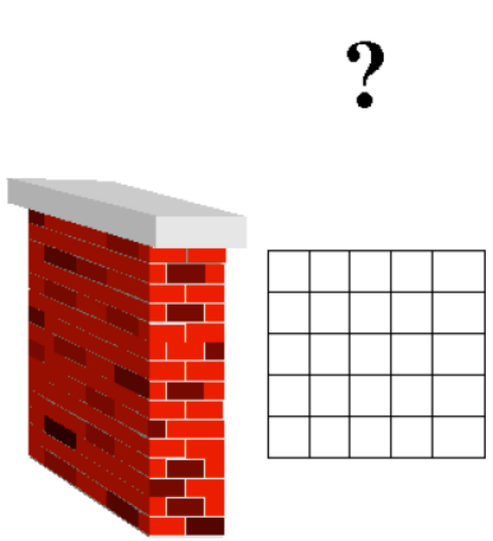
Bayes Decision Theory

Concept 1: **Priors** (a priori probabilities)

$$p(C_k)$$

- What we can tell about the probability before seeing the data.
- Example:





$$C_1 = a \qquad p(C_1) = 0.75$$

$$C_2 = b \qquad p(C_2) = 0.25$$

In general: $\sum_k p(C_k) = 1$

Bayes Decision Theory

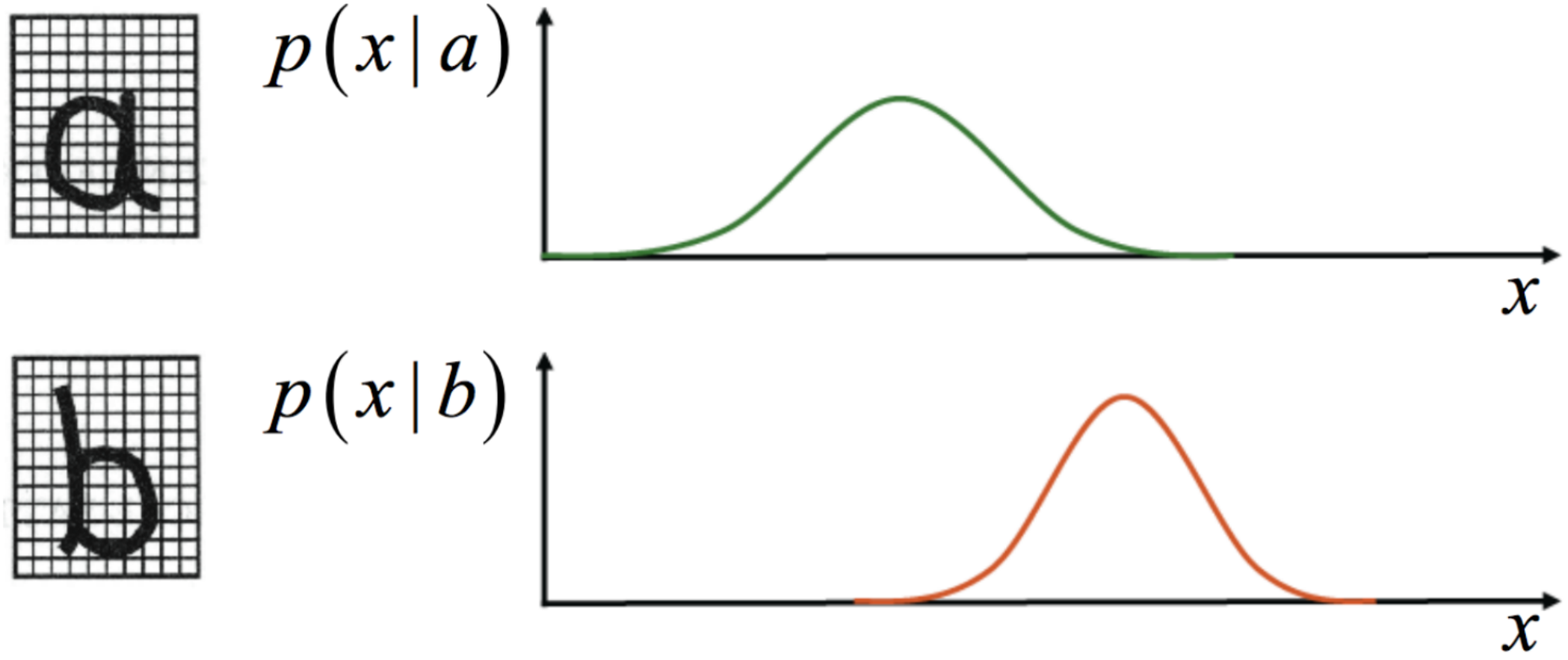
Concept 2: **Conditional probabilities**

$p(x|C_k)$

- Let x be a feature vector.
- x measures/describes certain properties of the input.

–E.g. number of black pixels, aspect ratio, …

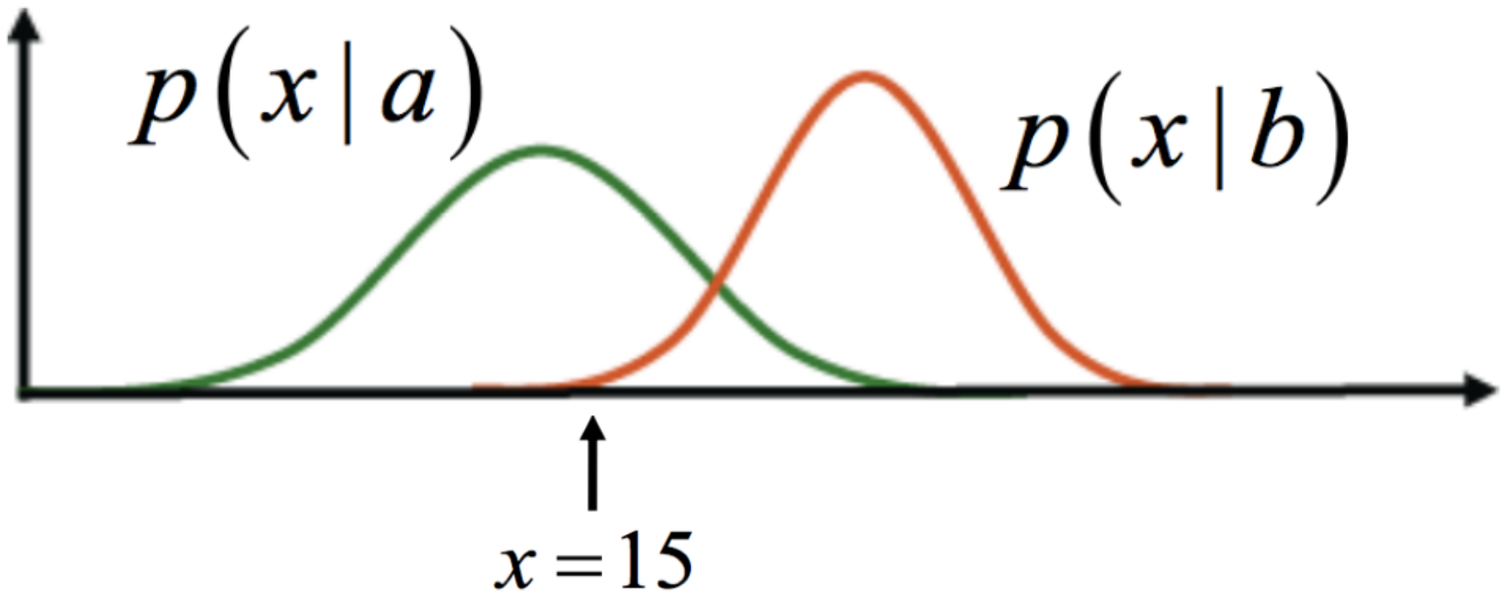
- $p(x|C_k)$ describes its **likelihood** for class C_k.



Slide source: B. Leibe Slides ML, 2015

Bayes Decision Theory

Example:



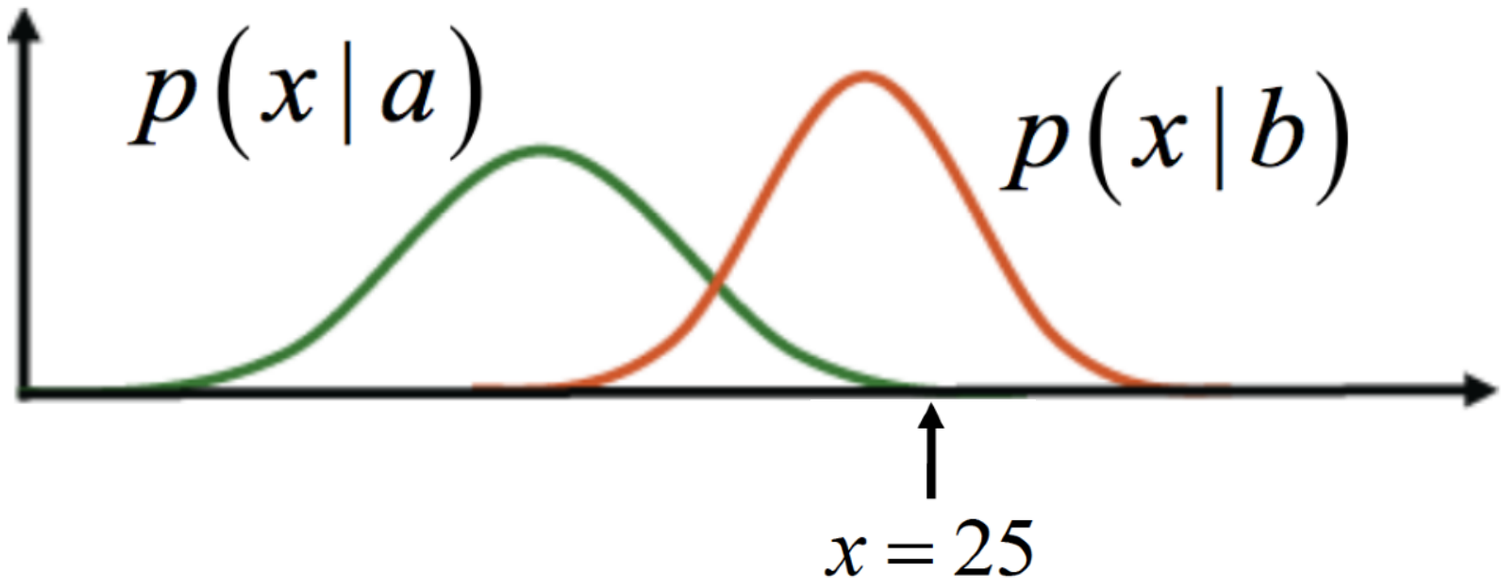
Question:

- Which class?
- Since $p(x|b)$ is much smaller than $p(x|a)$, the decision should be ‘a’ here.

Slide source: B. Leibe Slides ML, 2015

Bayes Decision Theory

Example:



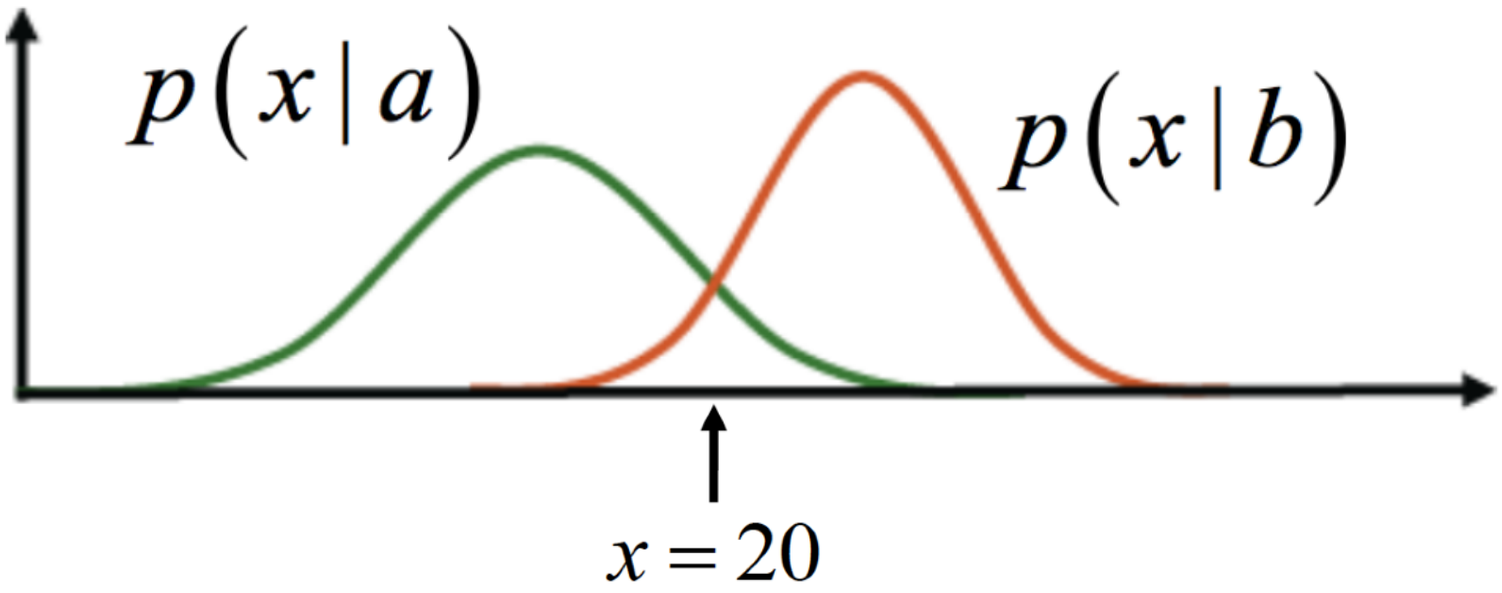
Question:

- Which class?
- Since $p(x|a)$ is much smaller than $p(x|b)$, the decision should be '*b*' here.

Slide source: B. Leibe Slides ML, 2015

Bayes Decision Theory

Example:



Question:

- Which class?
- Remember that $p(a) = 0.75$ and $p(b) = 0.25$…
- I.e., the decision should be again 'a'.

  => How can we formalize this?.

Slide source: B. Leibe Slides ML, 2015

Bayes Decision Theory

Concept 3: **Posterior probabilities**

$$p(C_k \mid x)$$

- We are typically interested in the a *posteriori* probability, i.e. the probability of class C_k given the measurement vector x

Bayes' Theorem:

$$p(C_k \mid x) = \frac{p(x \mid C_k)\, p(C_k)}{p(x)} = \frac{p(x \mid C_k)\, p(C_k)}{\sum_i p(x \mid C_i)\, p(C_i)}$$

Slide source: B. Leibe Slides ML, 2015

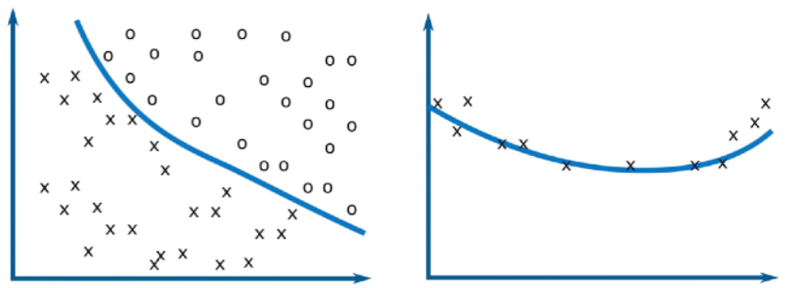
End of Video 1.5

Summary for Introduction Lecture

Summary

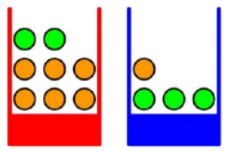
Machines that **learn** to **perform** a **task** from **experience**

Classification vs. Regression

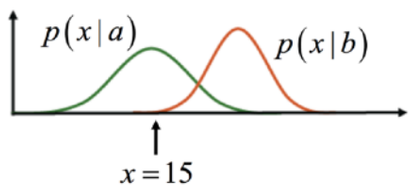


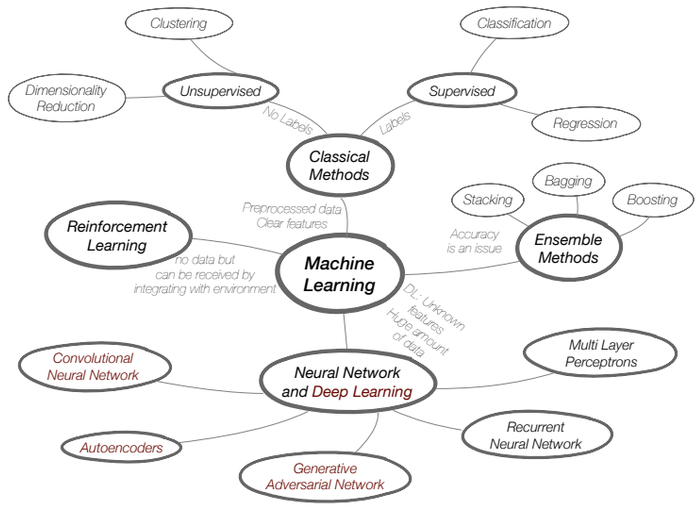
Type of data -> type of ML algorithm

Probability theory



Bayes Decision Theory





Next Lecture

Bayes Decision Theory

- Basic concepts
- Minimizing the misclassification rate
- Minimizing the expected loss

Probability Density Estimation

- General concepts
- Gaussian distribution

Parametric Methods

- Maximum Likelihood approach
- Bayesian vs. Frequentist view on probability
- Bayesian Learning

Readings

- More information, including a short review of Probability theory and a good introduction in Bayes Decision Theory can be found in Chapters 1.1, 1.2 and 1.5 of