Introduction to Learning and Probabilistic Reasoning

Dr. Francis Colas
Information Processing for Robotics
30.09.2011

Autonomous Systems Lab
ETH Zürich

Introduction

- Robotics:
  - Mechanics (moving parts, structure),
  - Electronics (motorize, power, control),
  - Software (OS, drivers, and the rest);

- To perform robotic tasks:
  - Make sense of sensor data
  - Decide on motor commands

=> information processing

Examples

- Robocup: robots playing football
  - Find where the ball is and how it's moving,
  - Know where you are in the field,
  - Know where your teammates are and what they are doing,
  - Same for opponents,
  - Plan some strategy,
  - Move and react to changing conditions,
  - Kick...

Examples

- Robocup:
  - Ball: color segmentation (which colors?);
  - Localization: field line extraction, beacons;
  - Teammates: communication;
  - Opponents: extraction of opponents, inference on actions;
  - Planning: prediction (using experience);
  - Moving, kicking: trained control.

Summary

- Coping with ignorance:
  - Things that we don't know,
  - Things that we don't know how to do;

- 2 set of tools:
  - Machine learning: adapt to experience,
  - Probabilistic reasoning: reason with uncertainties.

Learning: introduction

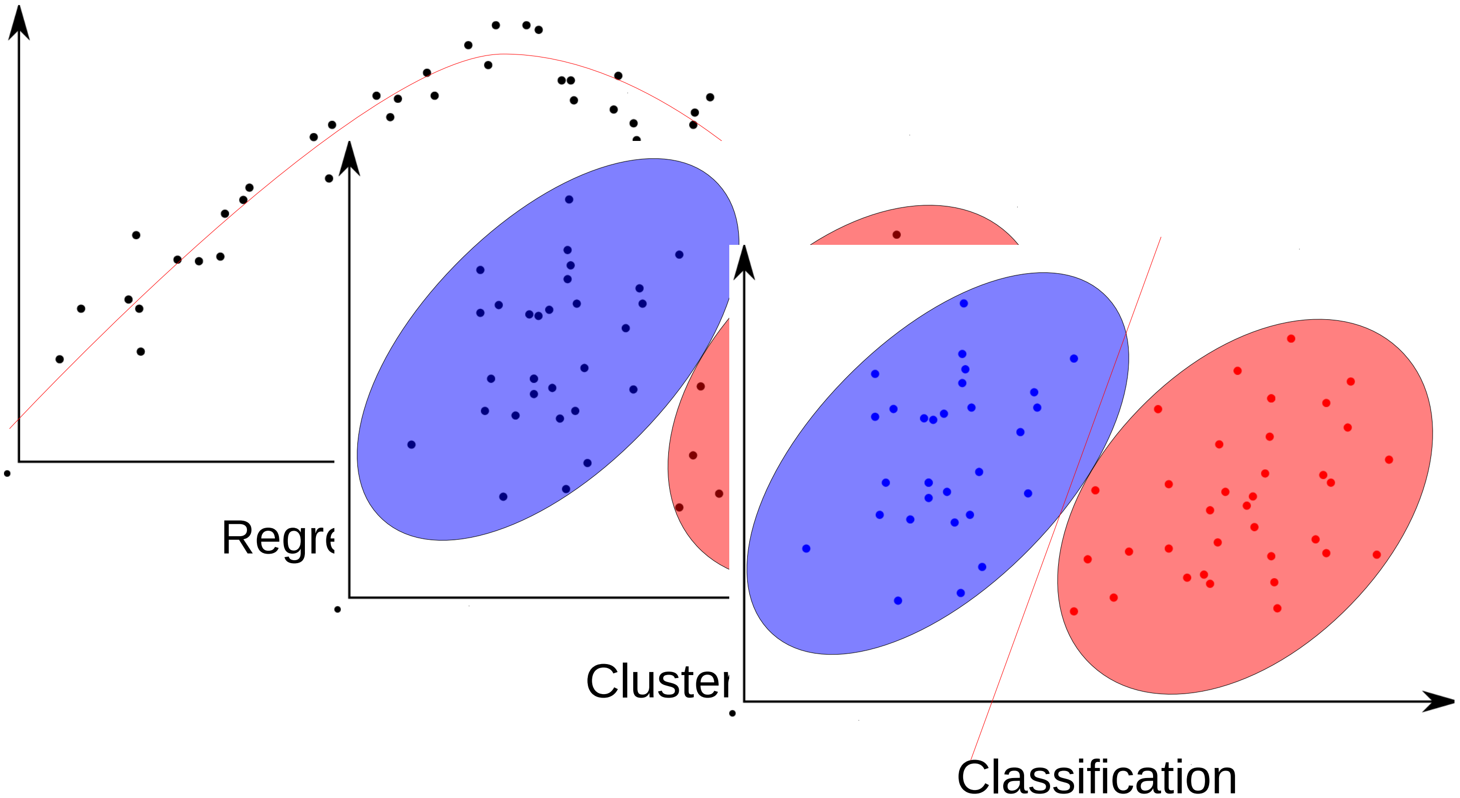
- Easy to sketch algorithms
  - e.g. color segmentation
- Difficult to tune to real conditions
  - Which colors, which threshold?

- Machine learning:
  - Adapt algorithms to empirical data,
  - Different things can be learned,
  - Different ways of learning,
  - Evaluation of learning.

What can we learn?

- Several things you can learn:
  - Simple parameters
    - e.g. color threshold
  - Regression analysis
    - Relationship between variables, curve fitting
    - e.g. finding ball trajectory
  - Clustering or classification of data
    - Separate complex data into several groups
    - e.g. teammate of opponent? ball or background?

What can we learn?



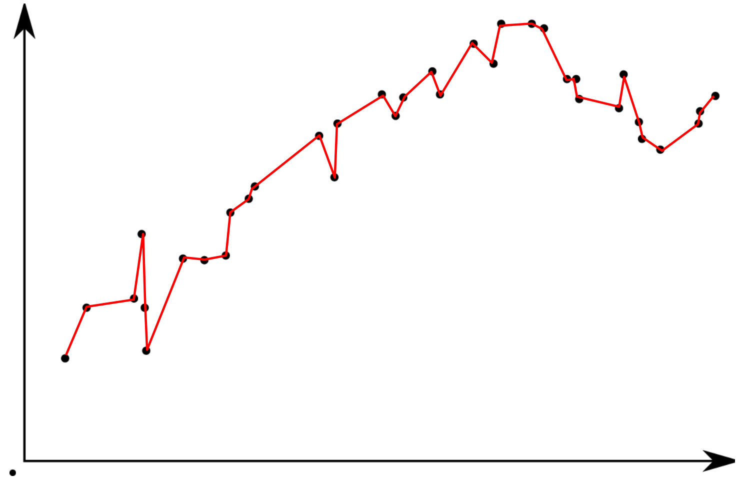
Regression

- Problems:
  - Find the best curve to fit the data,
  - Predict the value for a new data point;

- Formulation:
  - Let $\boldsymbol{X}$ be a set of points,
  - Let $\boldsymbol{y}$ be their corresponding value,
  - Find f such as $f(\boldsymbol{X}) \approx \boldsymbol{y}$
  - Find y for a new X,
  - Evaluate using goodness of fit.

Regression

- Overfitting:
  - Going through all points is good
  - But bad generalization
  - Complex model



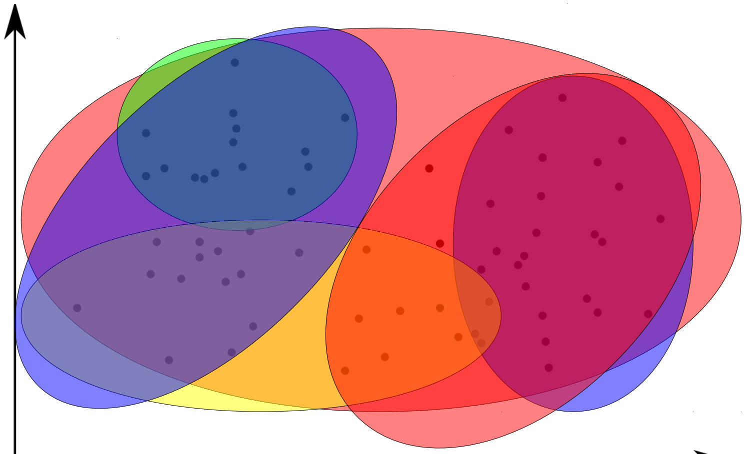
- Techniques:
  - Linear regression (class 5),
  - Gaussian processes (class 6).

Clustering

- Problems:
  - Group points into unknown classes,
  - Predict the class for a new data point;
- Formulation:
  - Let X be a set of points,
  - Define a set K of classes,
  - Find the association between X and k.

Clustering

- Ill-posed problem:
  - How many clusters?
  - Shape of clusters?



- Techniques:
  - *k*-means (class 10),
  - Expectation Maximization (class 11).

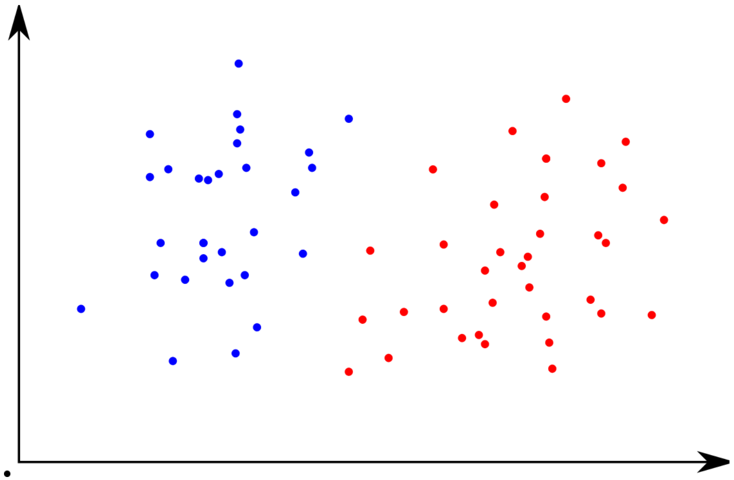
Classification

- Problems:
  - Group points into known classes,
  - Predict the class for a new data point;

- Formulation:
  - Let $\boldsymbol{X}$ be a set of points,
  - Let $\boldsymbol{k}$ be their respective class (or label),
  - Find the association between $\boldsymbol{X}$ and $\boldsymbol{k}$,
  - Find k for a new X,
  - Evaluate using rate of classification.

Classification

- Difference with clustering
  - Classes are known
  - Association is known



- Techniques:
  - Support Vector Machines (class 6),
  - Principal Components Analysis (class 13).

Different ways to learn

- Supervised learning:
  - Labels or target values known
  - e.g. regression, classification;
- Unsupervised learning:
  - No labels
  - e.g. clustering;
- Reinforcement learning:
  - Target value unknown
  - Reward or feedback given.

Evaluating learning

- Comparing different models:
  - Basis function for regression,
  - Shapes of classes for classification;

- *Cross-validation*
  - Partition the data set
  - Optimize on the *training data*,
  - Evaluate on the *test data*,
  - You can do that several times by changing partitions.

Summary on learning

- Learn different things:
  - Relationship between variables,
  - Clusters,
  - Classes...
- Different ways:
  - Supervised,
  - Unsupervised...
- Be careful with results:
  - Overfitting,
  - Cross-validation;
- Extracting knowledge from data.

Probabilistic reasoning

- Issues:
  - Sensor may fail,
  - Models are inaccurate,
  - Unexpected things happen,
  - =>Several sources of uncertainty;
- Represent uncertainty as probability
  - Probability values for different possibilities,
  - Reasoning by probabilistic computations;

Probabilistic reasoning

- Variables:
  - Relevant objects or quantities,
- Probability distributions:
  - Summarize the uncertainty on the values of variables,
- Relationship between variables:
  - Joint probability distribution,
  - Conditional probability distribution,
  - Independence;
- Inference rules:
  - Compute the (conditional) distribution over some variables based on some other distributions

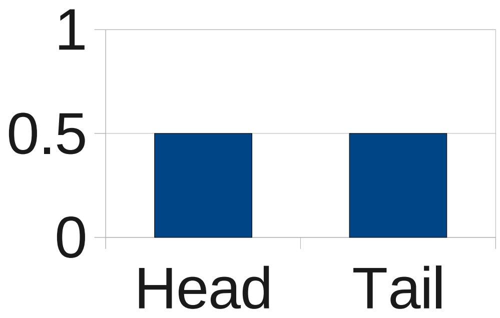
Variables

- Events, proposition, values...
- Examples:
  - result of a coin toss: $C \in \{Head, Tail\}$
  - dice outcome: $D \in \{1, 2, 3, 4, 5, 6\}$
  - distance to a beacon: $D \in \mathbb{R}^+$
  - pose: $P = (x, y, \theta) \in \mathbb{R} \times \mathbb{R} \times [0; 2\pi]$
  - ...
- Domain can be discrete or continuous
- Can be vectors
- Can be conjunction of variable
- Can be mixed

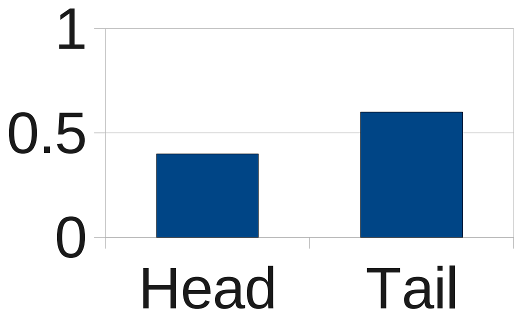
Probability distributions

- Discrete variables:
  - Coin:

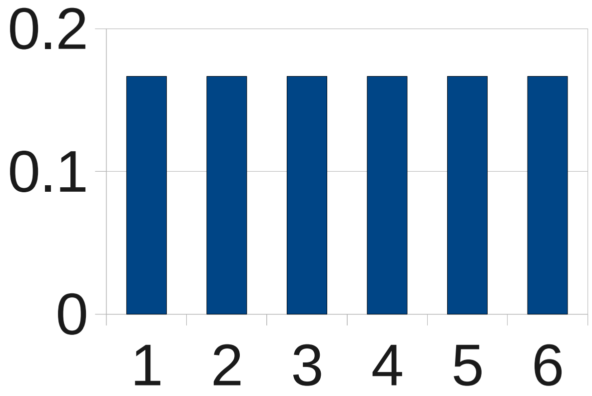
$P(Head)=0.5; P(Tail)=0.5$

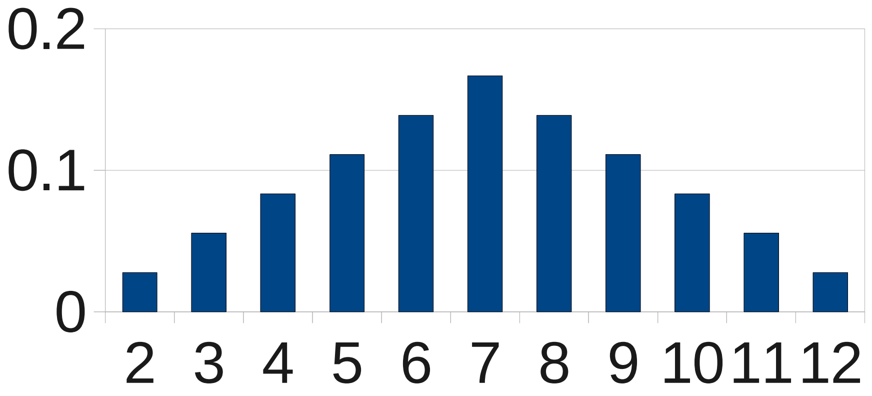


$P(Head)=0.4; P(Tail)=0.6$



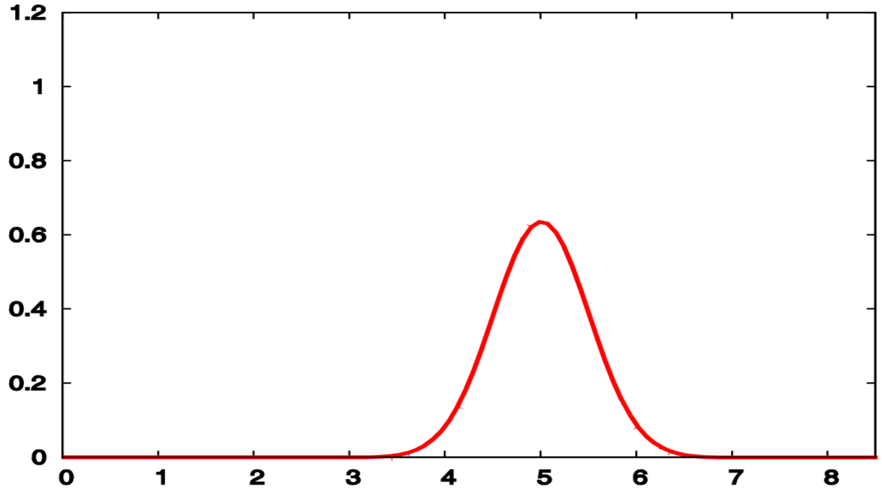
  - Dice:





Probability distributions

- Continuous: density function
  - Gaussian: $$x \in \mathbb{R},\ p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$$



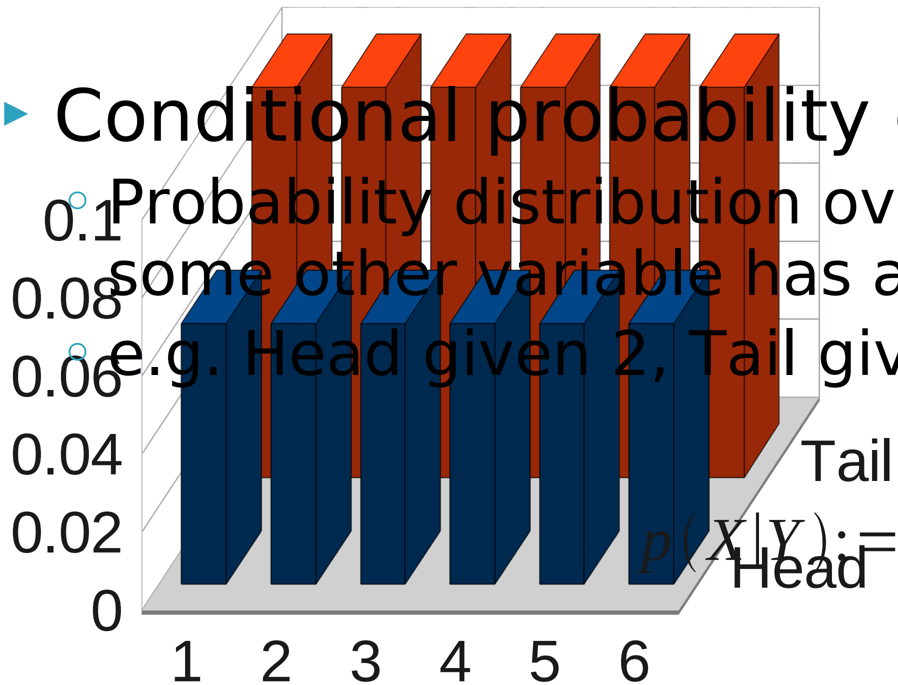
  - Multivariate Gaussian (on a vector)
  - Beta or Dirichlet
  - Exponential
  - ...

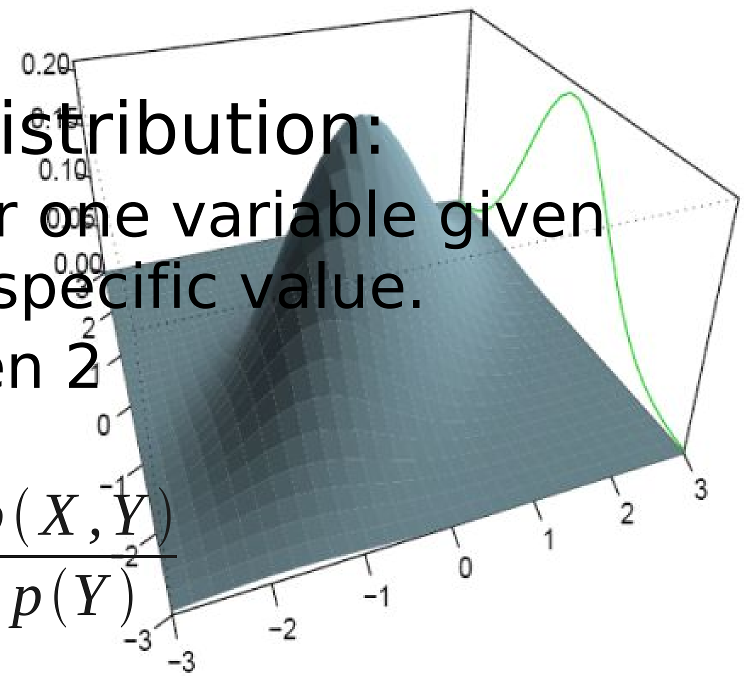
ETH Zürich

Relationship between variables

- Joint probability:
  - Probability of both variables having specific values:
  - e.g. Head and 1, Tail and 1, Head and 2, Tail and 2...

$$p(X \wedge Y) \quad p(X, Y)$$





- Conditional probability distribution:
  - Probability distribution over one variable given some other variable has a specific value.
  - e.g. Head given 2, Tail given 2

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

ETH Zürich

Relationship between variables

- Independence:
  - Value of X does not give information on Y,
  - X and Y are independent iff: $p(X,Y)=p(X)p(Y)$
  - e.g.: coin and dice;

- Conditional independence:
  - Given the value of Z, the value of X does not give information on Y,
  - X and Y are independent given Z iff:

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

  - Equivalent to:

$$p(X|Z)=p(X|Y,Z)\wedge p(Y|Z)=p(Y|X,Z)$$

Complexity

- 2 variables:
  - $P(A, B)$: distribution over the Cartesian product of A and B,
  - In case of independence: $P(A,B) = P(A)P(B)$
    - Distribution over A
    - Distribution over B
- 3 variables:
  - Conditional independence:
    - Distribution over A $\quad P(A,B,C) = P(A)P(B|A)P(C|A$
    - Conditional distribution over B given A
    - Conditional distribution over C given A
- (Cond.) Independence reduces complexity

Inference rules

- Sum rule:
  - Law of total probability,
  - Normalization of probability distributions:

$$P(A)=\sum_B P(A,B)$$

- Product rule:
  - Bayes' theorem,
  - From joint to conditional:

$$P(A,B)=P(A|B)P(B)$$

Inference rules

- We can deduce:

$$P(A,B)=P(A|B)P(B)=P(B|A)P(A)$$

$$P(B)=\sum_A P(A,B)$$

$$P(B)=\sum_A P(B|A)P(A)$$

$$P(A|B)=\frac{P(B|A)P(A)}{P(B)}$$

- General inference:

$$P(S|K)=\frac{\sum_F P(S,F,K)}{\sum_{S,F} P(S,F,K)}$$

Inference

- General inference:
  - Joint distribution over all variables,
  - Let S be the subset of variables you want,
  - Let K be the subset of variables whose value you know,
  - Let F be the rest of the variables,
  - Then:

$$P(S|K) = \frac{\sum_{F} P(S,F,K)}{\sum_{S,F} P(S,F,K)} \propto \sum_{F} P(S,F,K)$$

- Problems:
  - Specify the joint probability distribution,
  - High complexity in high dimensional space.

Example

- Noisy sensor:
  - Door detector
  - Specify if there is a door or not: S
  - 20% chance to not see the door and 10% chance to hallucinate it:

$P(S \mid D)$	S=True	S=False
Door	0.8	0.2
No door	0.1	0.9

  - A priori, 60% chance there is a door: $P(D=True)=0.6$
  - Sensor says no door, is there one or not?

$$P(D|S=False)?$$

$$P(D|S=False)=\begin{pmatrix} P(D=True|S=False) \\ P(D=False|S=False) \end{pmatrix}=\begin{pmatrix} \frac{0.2*0.6}{0.2*0.6+0.9*0.4} \\ \frac{0.9*0.4}{0.2*0.6+0.9*0.4} \end{pmatrix}=\begin{pmatrix} 0.25 \\ 0.75 \end{pmatrix}$$

Example

- Sensor fusion:
  - Adding a second sensor, T:

$P(T \mid D)$	T=True	T=False
Door	0.95	0.05
No door	0.05	0.95

  - Naive fusion: $P(D,S,T)=P(D)P(S|D)P(T|D)$
  - If they both see a door:

$$P(D|S,T)=\left(\begin{array}{c}\frac{0.6*0.8*0.95}{0.6*0.8*0.95+0.4*0.1*0.05}\\ \frac{0.4*0.1*0.05}{0.6*0.8*0.95+0.4*0.1*0.05}\end{array}\right)=\left(\begin{array}{c}0.996\\ 0.004\end{array}\right)$$

  - More certainty than any of the sensors.

Summary: Probabilistic Reasoning

- Aim:
  - Transform uncertainty into probability;
- Reasoning:
  - Specify the joint distribution,
  - Reduce complexity with (cond.) independence
  - General inference;
- Properties
  - Combine uncertain knowledge,
  - Fusion can reduce uncertainty;
- Difficulties
  - Computational complexity,
  - Specification of joint.

Summary

- Two techniques to cope with ignorance:
- Learning:
  - Adapt algorithm to empirical data,
  - Regression,
  - Clustering,
  - Classification;
- Probabilistic reasoning:
  - Cope with inherent uncertainty,
  - Inference.