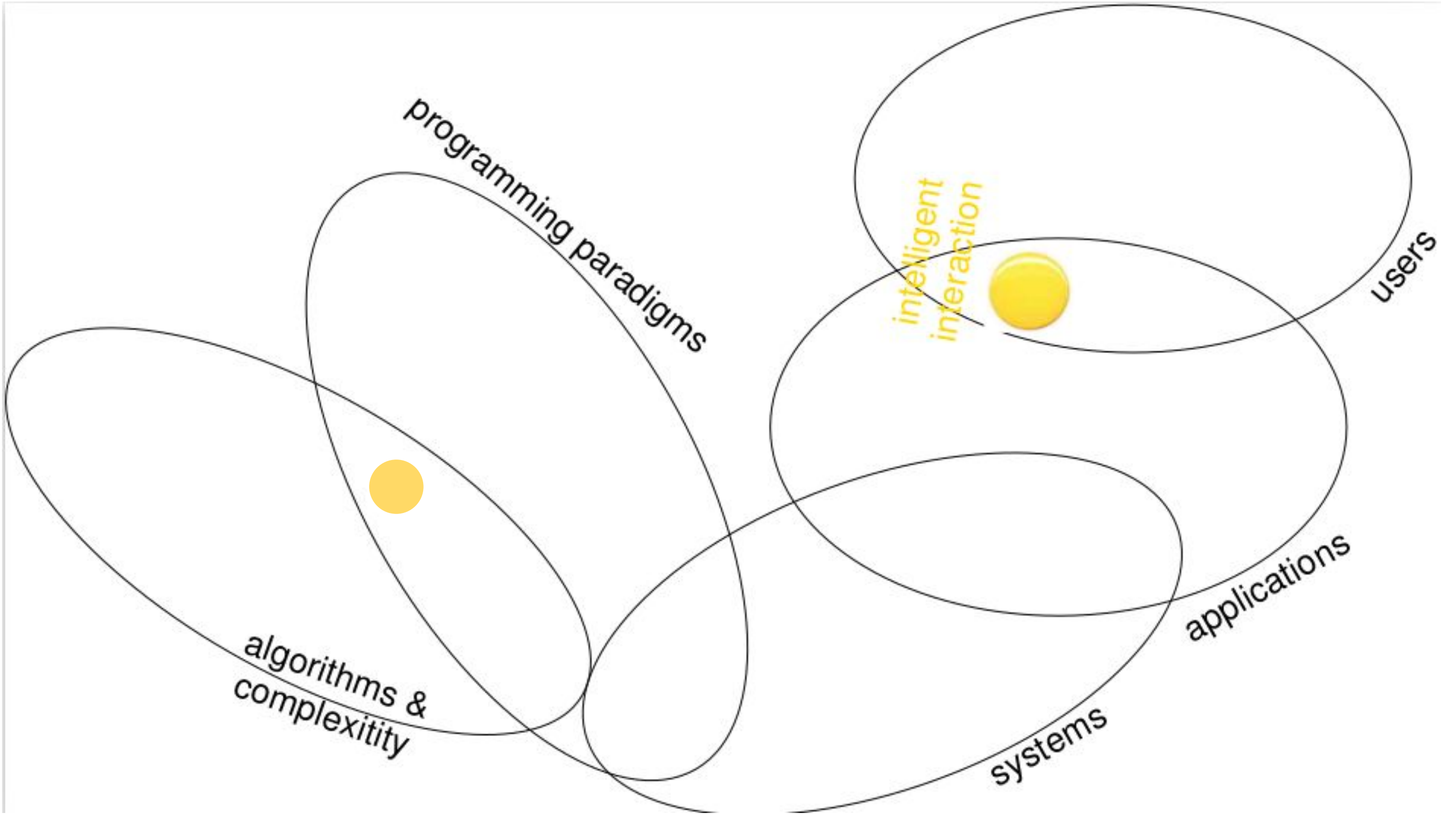
Pearl 110 - Intelligent Interaction

Introduction to Machine Learning
Bayesian Decision Theory

Intelligent Interaction Pearl



Contents

Intelligent Interaction

Machines that interact with humans or other machines, in an **intelligent way**

Examples: Robots, AI, avatars and VR

Intelligent interaction is difficult: What is intelligence?

...for a machine:

Hard to predict all possible answers/actions and plan ahead

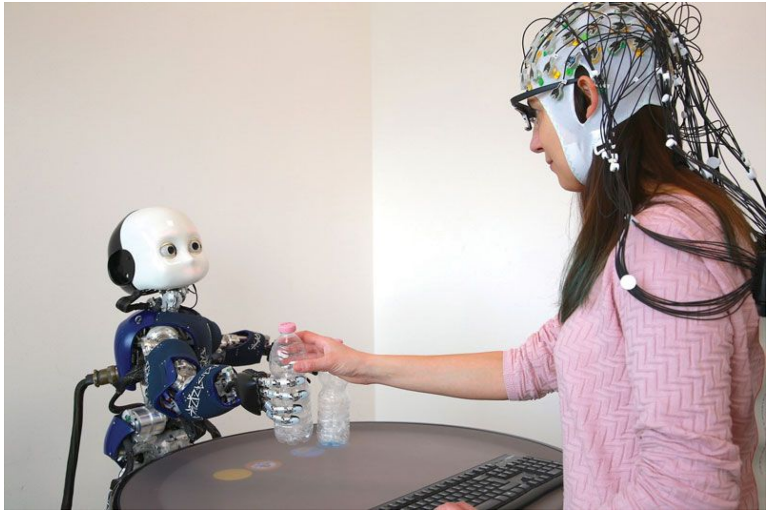
There might be unobserved variables (user history, expectations, stress, etc.)

Intelligent Interaction

Sensing the world
cameras/microphones
tactile/pressure sensors
proximity sensors
...

Understanding the world
computer vision/sound analysis
space perception
emotion recognition
(face/speech)
…

Acting with the world
plan actions/take them
consider feedback



Humans are (usually) **good at interacting** with or 'reading' other humans

Recognize emotion and act accordingly

Get tired (especially on repeated tasks)

Machine Learning and Intelligent Systems

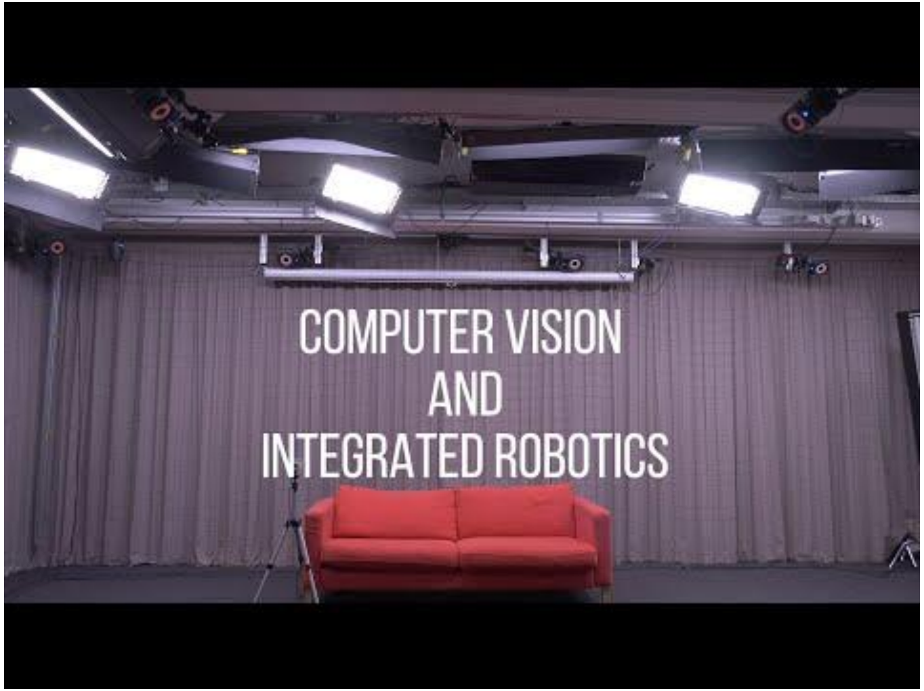
We make machines perceive and understand the world, and take decisions upon what is perceived (reacting on inputs)

What? Algorithms and techniques to get machines to learn from examples

Why? We can't do a task ourselves (but we know it should be possible)
...or *we don't know how* to do a task

Music recognition: song identification

Robotics (self-driving cars)



Robotics and Computer Vision (TrimBot2020, www.trimbot2020.org)

Organization: general

3 Lectures

Monday, 13:45 -15:30 (Zoom)

Tuesday, 10:45 - 12:30 (Zoom)

Wednesday, 13:45 - 15:30 (Zoom)

Q&A session

Thursday , 11:45 - 12:30 (Zoom) [use the form (on canvas) to submit questions]

Study materials and information

See Canvas

Exam

Friday 21st, 8:45 - 10:45

Organization: Pearl assignment

3 assignments

See the pearl guide - Assignments 7.5, 7.6, 7.7

1 bonus assignment

Assignment 7.8 - you will need to program in a Python notebook

It is worth a +1 in the grade (0 or 1)

Work in groups of size 2

Deadlines

Bonus assignment: <u>Friday at 23:59</u> (chance of getting the bonus)

Pearl assignments: <u>Sunday at 23:59</u> (no bonus)

Organization: tutorials

No sign-off lists

Tutorial sessions: you get help from TAs

Monday, 15:45-17:30

Tuesday, 13:45-15:30

Thursday, 8:45- 12:30

Tutorial assignment solutions: presentations from TAs

Monday, ~16:30 (in the lecture rooms)

Tuesday, ~14:15 (in the lecture rooms)

Thursday, ~10:00 (in the lecture rooms)

Organization of the week

Mon 17 Oct	Tue 18 Oct	Wed 19 Oct	Thu 20 Oct	Fri 21 Oct
08:45 - 10:30 **Introduction to Mathematics + Calculus 1A for CS 202001190** Other Online *Diagnostic Test & Feedback* *Student set(s): CS MOD01 Group 01, CS MOD01 Group 02, CS MOD01 Group 03, CS MOD01 Group 04, MI CS MOD01 Group 01*	**09:45 - 10:30** **Pearl 202001022** Other	**08:45 - 10:30** **Introduction to Mathematics + Calculus 1A for CS 202001190** Self study supervised **08:45 - 10:30** **Introduction to Mathematics + Calculus 1A for CS 202001190** Self study supervised **08:45 - 10:30** **Introduction to Mathematics + Calculus 1A for CS 202001190** Self study supervised **08:45 - 10:30** **Introduction to Mathematics + Calculus 1A for CS 202001190** Self study supervised	**08:45 - 11:30** **Pearl 202001022** Practical *By TAs* SP 2 SP 3 SP 4 *Student set(s): AM/CS MOD01 Group 01 (double degree), CS MOD01 Group 01, CS MOD01 Group 02, CS MOD01 Group 03, CS MOD01 Group 04, and 4 other...*	**08:45 - 09:45** **Pearl 202001022** Exam *Test 6* **08:45 - 09:45** **Pearl 202001022** Exam *Test 6* **08:45 - 09:45** **Pearl 202001022** Exam *Test 6* **09:45 - 10:45** **Pearl 202001022** Exam *Test 7* **09:45 - 10:45** **Pearl 202001022** Exam *Test 7* **09:45 - 10:45** **Pearl 202001022** Exam *Test 7*
10:45 - 12:30 **Introduction to Mathematics + Calculus** Lecture Recorded F Kula T Akkaya *Student set(s): BIT MOD01 Group 01, BIT MOD01 Group 02, BMT MOD01 groep 01, BMT MOD01 groep 02, BMT MOD01 groep 03, and 92 other...*	**10:45 - 12:30** **Pearl 202001022** Lecture N Strisciuglio WA 2 *Student set(s): AM/CS MOD01 Group 01 (double degree), CS MOD01 Group 01, CS MOD01 Group 02, CS MOD01 Group 03, CS MOD01 Group 04, and 4 other...*	**10:45 - 12:30** **Introduction to Mathematics + Calculus 1A for CS 202001190** Self study unsupervised *Student set(s): CS MOD01 Group 01, CS MOD01 Group 02, CS MOD01 Group 03, CS MOD01 Group 04, MI CS MOD01 Group 01, and 3 other...*	**11:45 - 12:30** **Pearl 202001022** Q&A Online	
13:45 - 15:30 **Pearl 202001022** Lecture N Strisciuglio WA 2 *Student set(s): AM/CS MOD01 Group 01 (double degree), CS MOD01 Group 01, CS MOD01 Group 02, CS MOD01 Group 03, CS MOD01 Group 04, and 4 other...*	**13:45 - 15:30** **Pearl 202001022** Practical *By TAs* SP 1 SP 3 SP 4 *Student set(s): AM/CS MOD01 Group 01 (double degree), CS*	**13:45 - 15:30** **Pearl 202001022** Lecture N Strisciuglio WA 1 *Student set(s): AM/CS MOD01 Group 01 (double degree), CS MOD01 Group 01, CS MOD01 Group 02, CS MOD01 Group 03, CS MOD01 Group 04, and 4 other...*	**13:45 - 15:30** **Introduction to Mathematics + Calculus 1A for CS 202001190** Tutorial RFJ van Lingen SP 3 **13:45 - 15:30** **Introduction to Mathematics + Calculus 1A for CS 202001190** Tutorial A Velicka MC Carioni **13:45 - 15:30** **Introduction to Mathematics + Calculus 1A for CS 202001190** Tutorial M Walter S Langer	
15:45 - 17:30 **Pearl 202001022** Practical *By TAs* SP 2 SP 3 SP 4 *Student set(s): AM/CS MOD01 Group 01 (double degree), CS*	**15:45 - 17:30** **Pearl 202001022** Self study unsupervised *Student set(s): AM/CS MOD01 Group 01 (double degree), CS MOD01 Group 01, CS MOD01 Group 02, CS MOD01 Group 03, CS MOD01 Group 04, and 4 other...*	**15:45 - 17:30** **Pearl 202001022** Self study unsupervised *Student set(s): AM/CS MOD01 Group 01 (double degree), CS MOD01 Group 01, CS MOD01 Group 02, CS MOD01 Group 03, CS MOD01 Group 04, and 4 other...*	**15:45 - 17:30** **Introduction to Mathematics + Calculus 1A for CS 202001190** Self study unsupervised *Student set(s): CS MOD01 Group 01, CS MOD01 Group 02, CS MOD01 Group 03, CS MOD01 Group 04, MI CS MOD01 Group 01, and 3 other...*	**15:45 - 16:30** **Academic Skills 202001022** Self study unsupervised

Mathematics is fundamental for ML!

- 1st and 2nd order equations
- Binomial distribution
- Exponential, Logarithm, Gaussian functions
- Some basics of linear algebra
  - Vector arithmetics
  - Matrix multiplication (up to 2x2 matrices)

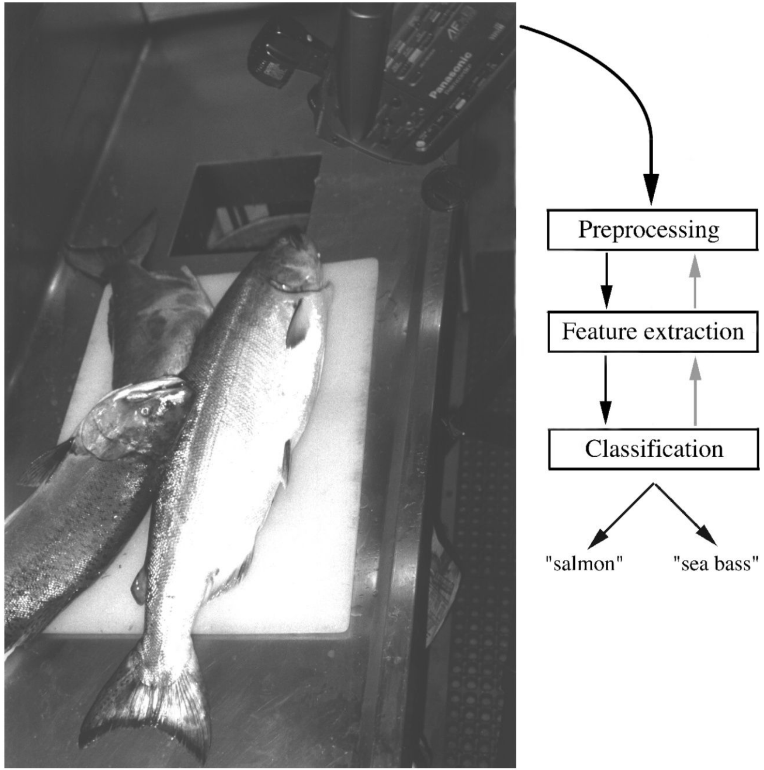
- … much more for Neural Networks, Deep Learning, Computer Vision…
  (not now, not here!)

How machines learn and decide

Recognizing natural objects

Example: automatic sorting of fishes on a conveyor belt, for automation of fish-can production

We need to **measure** some characteristics (*features*) of the fishes and use them to **decide** what actions to take

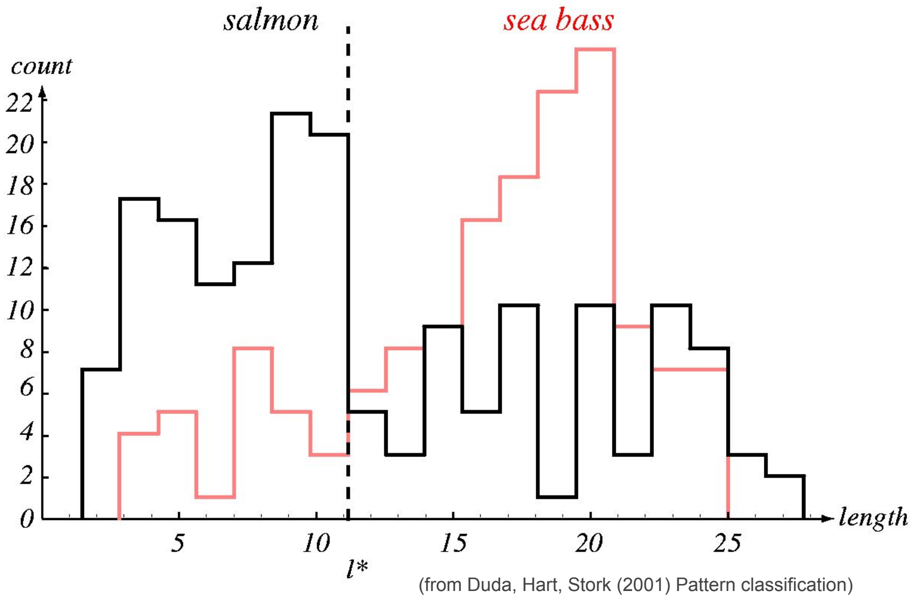


(from Duda, Hart, Stork (2001) Pattern classification)

Extract features from natural objects

Histogram of the measured **length** of a number of fishes (training samples) on a conveyor belt.

Use the histogram of the feature values to take decisions.

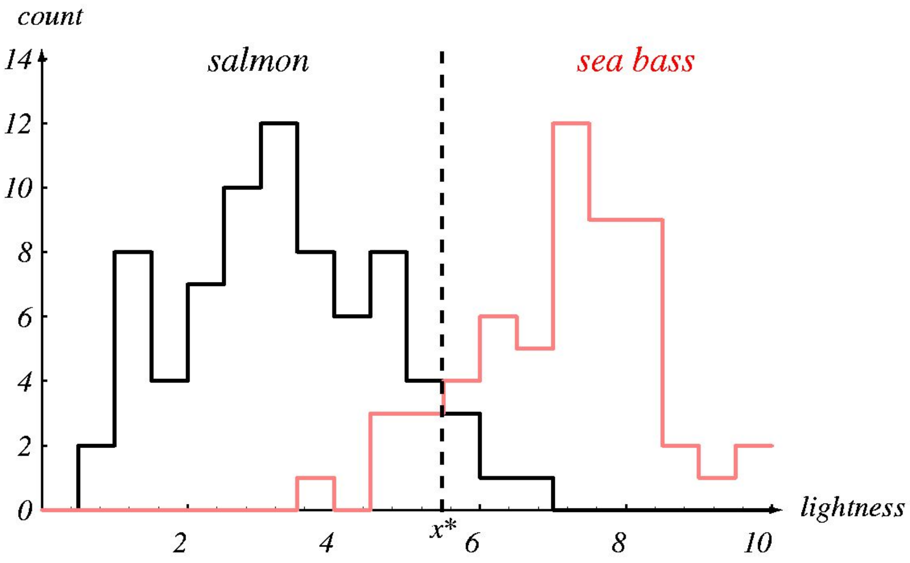


(from Duda, Hart, Stork (2001) Pattern classification)

Extract features from natural objects

Histogram of the measured **lightness** of a number of fishes (training samples) on a conveyor belt.

Use the histogram of the feature values to take decisions.



(from Duda, Hart, Stork (2001) Pattern classification)

Choose a value separate the samples (fishes)

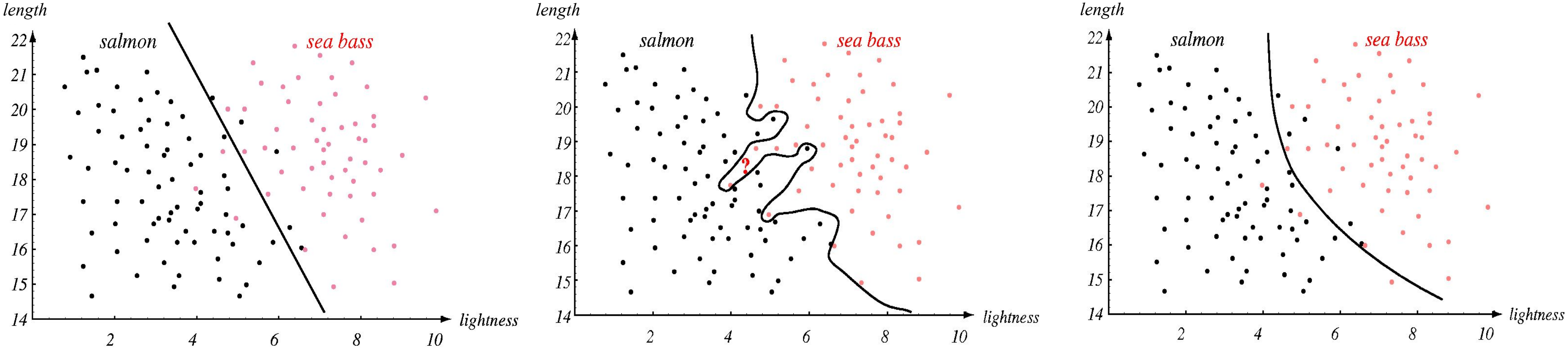
Classification is done by taking a certain value, called ***decision criterion***, and marking objects as belonging to class 1 (salmon) or class 2 (seabass), depending on whether the features are smaller or larger than the decision criterion.

One single feature (length or lightness) might not be enough to achieve good classification. We can pack them in a feature vector **x = [length, lightness]**.

In the case when *multiple features* are used for the decision, we use a **decision boundary**.

Possible boundaries

One can **learn** different decision boundaries (i.e. different ML models)



(from Duda, Hart, Stork (2001) Pattern classification)

Learning

A pattern (object) that is represented by a feature vector needs to be classified.

This can be achieved by comparing this feature vector with previously stored feature vectors for which it is known what type of object they represent.

The process of constructing a classification model with such feature vectors can be considered as '**learning**': using training data to automatically determine/adjust some inner parameters of a classifier system (e.g. the slope of a line, coefficients of a polynomial function, weights of convolutional kernels, ...).

Machine Learning

Supervised learning

Learning a mapping between input and output, using ground truth labels for training samples

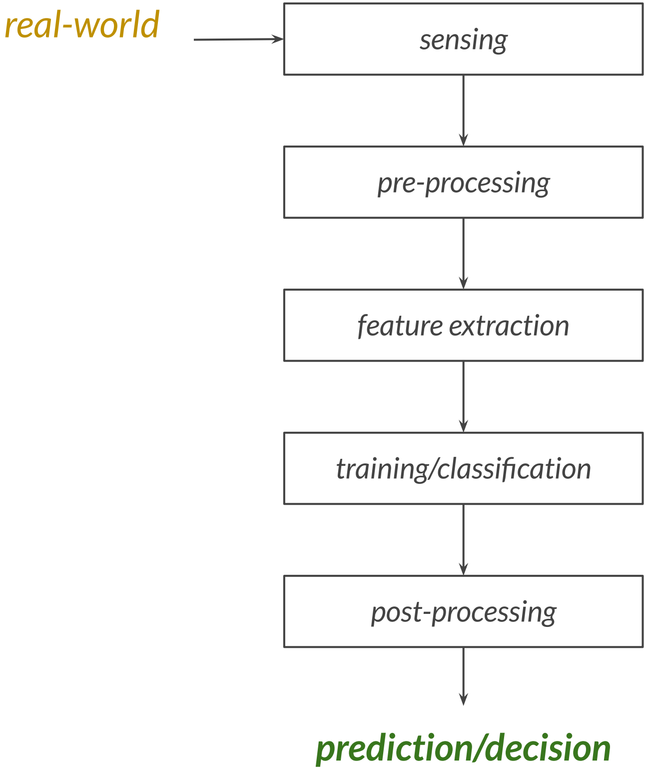
Unsupervised (self-supervised) learning

Learning a mapping between input and output without labels (e.g. grouping samples together by similarity)

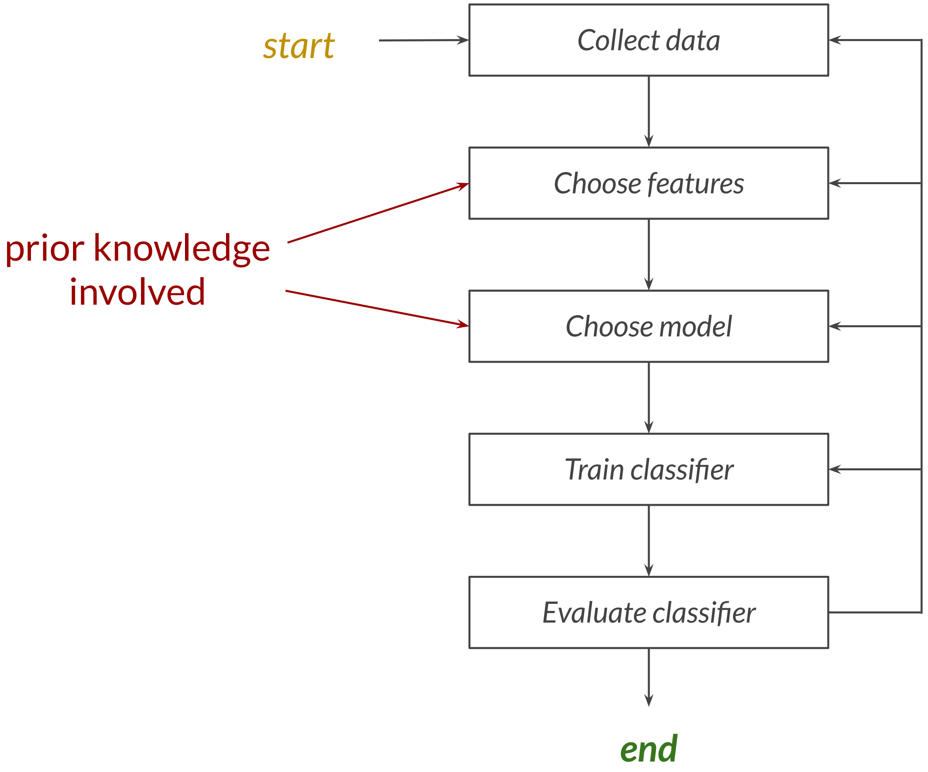
Reinforcement learning

Learning how software agents ought to take actions, optimizing a cumulative rewards.

Machine Learning systems: general structure



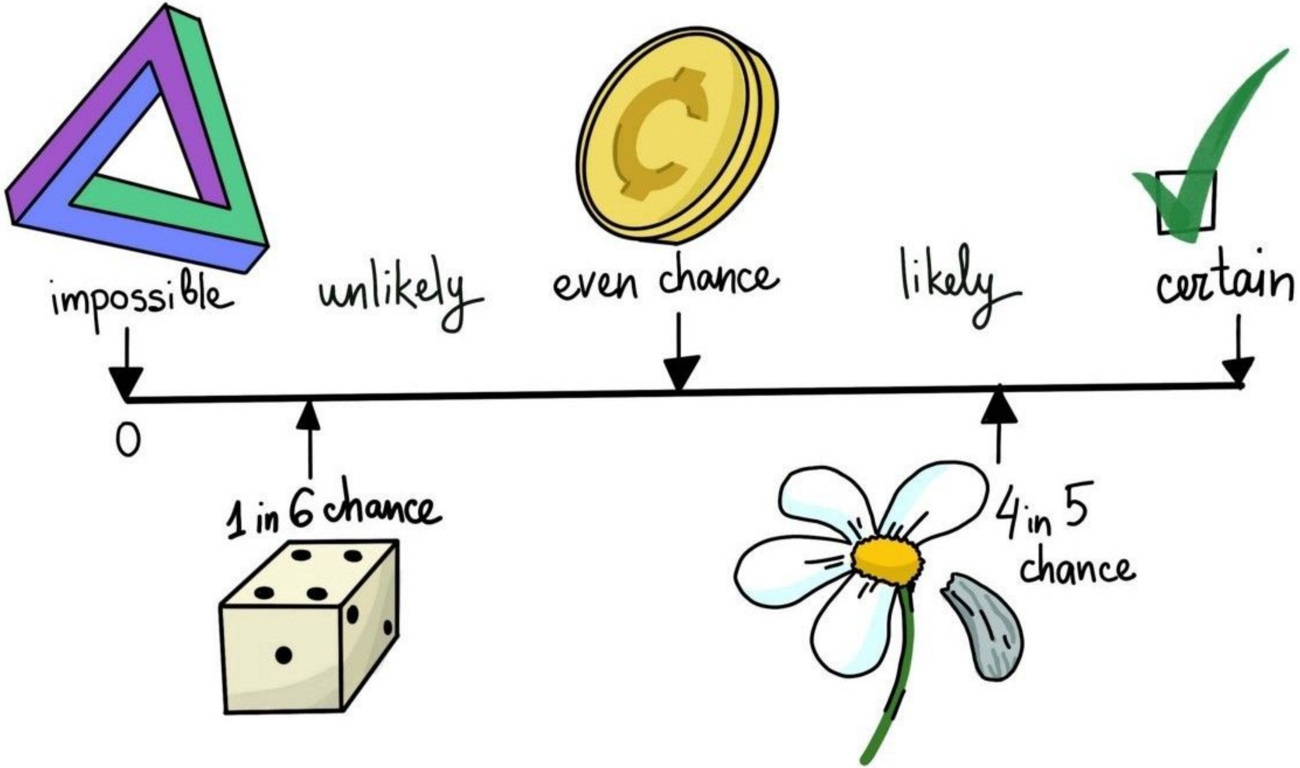
Machine Learning systems: design workflow



Introduction to Bayesian Decision Theory

A probabilistic approach to *intelligent* decisions

Dealing with probability of events



Event

We have a bucket with 10 fishes (4 salmon and 6 sea bass).
We do an experiment, and obtain an outcome.

The **outcome of the experiment is an event,** and can be:

- We pick a salmon from the bucket
- We pick a fish of width 26cm
- We pick a sea bass of 25cm
- ...

To each of these events we can assign a probability

Different types of event

Terminology that we use in probabilistic decision theory.

- **Prior**: probability of an event to happen, before (prior) to measuring any characteristics (features) of the data
- **Evidence:** probability of *measuring a certain value of features* from all the samples
- **Class-conditional:** probability of *measuring a certain feature value* from samples that *belong to a certain class* (conditioned on the class)
- **Posterior:** probability of a sample to *belong to a certain class*, *given that* we have measured *a certain feature value*

Prior probability

We have a bucket with 10 fishes (4 salmon and 6 sea bass).

Experiment: let us pick a fish from the bucket without looking.

We do not know anything about their characteristics (features): width, lightness, weight are unknown.

We only know that they are distributed as 4 salmon and 6 sea bass (prior information).

Without new information, we can only estimate that we have a chance of:

4 on 10 to pick a salmon (4/10 = 0.4) **P(salmon) = 0.4**

6 on 10 to pick a sea bass (6/10 = 0.6) **P(sea bass) = 0.6**

Evidence

Now we perform an experiment: we measure some characteristic of the fishes.

We measure the width of the fishes.
In probability terms we collect evidence.

The outcome is:

Salmon: *30cm, 25cm, 26cm, 26cm*

Sea bass: *24cm, 28cm, 26cm, 22cm, 24cm, 22cm*

What is the probability of finding the evidence *width=26cm*?

3 out of 10 fish have width 26cm (because we have evidence of it).
We can assign the probability to the event of measuring 26cm as 3 out of 10 samples:

P(width=26cm) = 3/10 = 0.3

Evidence probability: is the probability of measuring a certain value for a specific characteristic (feature) of our samples.

Class-conditional probability

We are **given an event**: a *fish from the class salmon*.

We perform an experiment: measure the width and obtain a certain outcome (e.g. width=26cm)

This time the <u>outcome</u> of the experiment (the event) <u>is conditioned on another event</u>

We can estimate the probability of having 26cm width as an outcome of our experiment, but this time it is conditioned on the event ‘a fish from the class salmon’.

What is the probability of measuring a width of 26cm, given that we have a salmon?

Probability of event 26cm

Conditioned by class salmon

P(width=26cm|salmon) = ?

Class-conditional probability (2)

What is the probability of measuring a width of 26cm, given that we have a salmon?

We have previously measured the width of the fishes:

Salmon: *30cm, 25cm, 26cm, 26cm*

Sea bass: *24cm, 28cm, 26cm, 22cm, 24cm, 22cm*

We are conditioned on the class salmon, so we only look at the width that we have available for samples of the class salmon

We see that we have 2 chances out of 4 to pick a salmon of width 26cm:

P(width=26cm|salmon) = 2/4 = 0.5

Class-conditional probability: we estimated the probability of the event 'measuring 26cm', conditioned by the event 'class salmon'.

Posterior probability

We have modeled our bucket of fishes, of which we know:

- **prior** probability (probability each class without measuring evidence)
- **evidence** (probability of measuring certain feature values)
- **class-conditional** probability (probability of measuring certain feature values, given that the samples belong to a certain class)

New experiment
We get a <u>new</u> fish, for which we measure the width (26cm):
this is <u>NEW</u> information, which is not in our data set (bucket).

We can compute the probability of it belonging to the class salmon, given the event of measuring width=26cm

It is called **posterior probability**, because you can compute it <u>AFTER</u> you have measured the new evidence (i.e. width of the new sample)

Posterior probability (2)

It is called **posterior probability**, because you can compute it AFTER you have measured the feature of the new sample

So far we have computed the probability of known events, because we have them in the data set

The posterior is the probability of an unknown event (the class of the new fish).

The fish is a <u>NEW</u> sample: it comes from the conveyor belt

We do not know its class

We can only measure its width

We thus compute the probability of assigning the class 'salmon' to the new fish, given that we measure its width

P(salmon| width = 26cm)

Posterior probability with the ***Bayes formula***

Joint probability
Probability of two events occurring at the same time
(e.g. measuring features **x** and class **C** of a sample)

$$P(x, C) = P(C|x)P(x)$$
$$P(x, C) = P(x|C)P(C)$$

$$P(C|x)P(x) = P(x|C)P(C)$$

class-conditional

$$\underbrace{\boxed{P(C|x)}}_{\text{posterior}} = \frac{\overbrace{\boxed{P(x|C)}}^{\text{class-conditional}}\,\overbrace{\boxed{P(C)}}^{\text{prior}}}{\underbrace{\boxed{P(x)}}_{\text{evidence}}}$$

The evidence is computed as: $P(X) = P(X|C_1)P(C_1) + P(X|C_2)P(C_2) + \ldots$

Posterior probability (3)

We thus compute the probability of assigning the class ‘salmon’ to the new fish, given that we measure its width.

Bayes formula

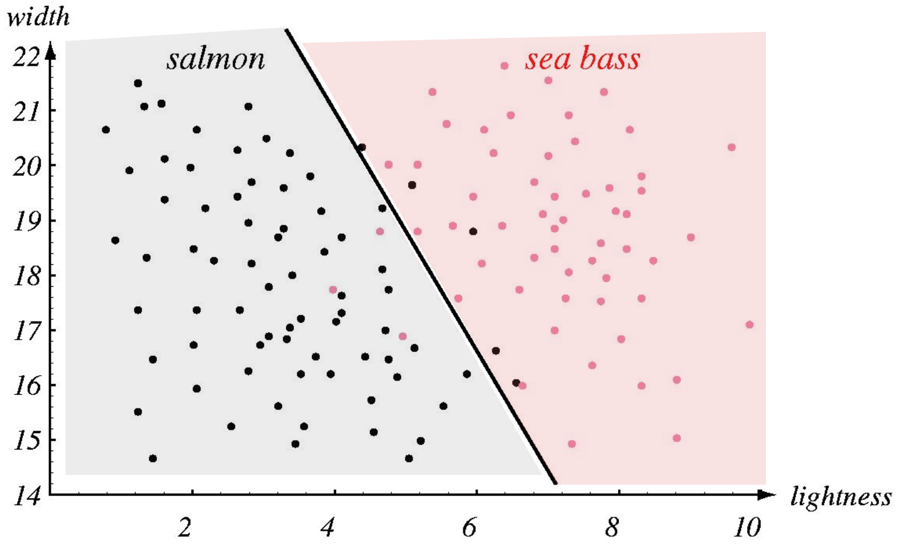
$$P(\text{salmon} \mid \text{width} = 26\text{cm}) = \frac{P(\text{width} = 26\text{cm} \mid \text{salmon})\ P(\text{salmon})}{P(\text{width}=26\text{cm})}$$

$$= \frac{0.5 * 0.4}{0.3} = 0.67$$

Posterior probability for classification

For each point, estimate the probability of belonging to each class.

<u>Choose the class with the highest posterior probability</u>



(from Duda, Hart, Stork (2001) Pattern classification)

Estimating probabilities

An example, and how to use them to make decisions

Estimation of probabilities: an example

For a shipping company, it is important to estimate how dangerous the combination of the shipped goods and type of the box chosen by the customer is (e.g. is it ok to ship wine in a thin cardboard package?). The customer care department recorded the following data:

object	package	OK shipping	Not OK shipping
wine	thin	3	7
plates	medium	5	8
plates	hard	8	3
wine	medium	5	4
	total:	**21**	**22**

Question: **Is it OK or not OK to ship plates in a medium package?**

Question: **Is it OK or not OK to ship plates in a medium package?**

What is the probability of having an **OK** (or a **not OK**) shipping, if we pack **plates** in a **medium cardboard**?

$$P(OK \mid plates, m) = ?$$

We can use the Bayes formula!

$$P(OK \mid plates, m) = \frac{P(plates,m|OK)P(OK)}{P(plates,m)}$$

object	package	OK shipping	Not OK shipping
wine	thin (t)	3	7
plates	medium (m)	5	8
plates	hard (h)	8	3
wine	medium (m)	5	4
	total:	**21**	**22**

Estimation of **prior** probability (example)

We only look at the distribution of the data, without measuring any features (no evidence). Can we estimate the probability of having an **OK** and a **Not OK** delivery?

object	**package**	**OK shipping**	**Not OK shipping**
wine	thin	3	7
plates	medium	5	8
plates	hard	8	3
wine	medium	5	4
	total:	**21**	**22**

Total shippings: 43

Estimation of **prior** probability (example)

We only look at the distribution of the data, without measuring any features (no evidence). Can we estimate the probability of having an **OK** and a **Not OK** delivery?

We do not look at the features (for now!)

object	package	OK shipping	Not OK shipping
[redacted]			
	total:	21	22

Total shippings: 43

Estimation of **prior** probability (example)

We only look at the distribution of the data, without measuring any features (no evidence). Can we estimate the probability of having an **OK** and a **Not OK** delivery?

We do not look at the features (for now!)

object	package	OK shipping	Not OK shipping
[redacted]			
	total:	21	22

Total shippings: 43

21 out of 43 shippings are **OK**

$$P(OK) = \frac{21}{43} \approx 0.488$$

22 out of 43 shippings are **Not OK**

$$P(\neg OK) = \frac{22}{43} \approx 0.512$$

Estimate **posterior** probability using Bayes

What is the probability of having an **OK** (or a **not OK**) shipping, if we pack **plates** in a **medium cardboard**?

$$P(OK \mid plates, m) = \frac{P(plates{,}m|OK)P(OK)}{P(plates{,}m)}$$

$$P(OK) = \frac{21}{43} \approx 0.488$$

object	package	OK shipping	Not OK shipping
wine	thin (t)	3	7
plates	medium (m)	5	8
plates	hard (h)	8	3
wine	medium (m)	5	4
	total:	**21**	**22**

Estimate **posterior** probability using Bayes

What is the probability of having an **OK** (or a **not OK**) shipping, if we pack **plates** in a **medium cardboard**?

$$P(OK \mid plates, m) = \frac{P(plates, m | OK) P(OK)}{P(plates, m)}$$

object	package	OK shipping	Not OK shipping
wine	thin (t)	3	7
plates	medium (m)	$\boxed{5}$	8
plates	hard (h)	8	3
wine	medium (m)	5	4
	total:	$\boxed{\mathbf{21}}$	**22**

$$P(OK) = \frac{21}{43} \approx 0.488$$

$$P(plates, m | OK) = \frac{5}{21} \approx 0.24$$

Estimate **posterior** probability using Bayes

What is the probability of having an **OK** (or a **not OK**) shipping, if we pack **plates** in a **medium cardboard**?

$$P(OK \mid plates, m) = \frac{P(plates,m|OK)P(OK)}{P(plates,m)}$$

object	package	OK shipping	Not OK shipping
wine	thin (t)	3	7
plates	medium (m)	5	8
plates	hard (h)	8	3
wine	medium (m)	5	4
	total:	**21**	**22**

$$P(OK) = \frac{21}{43} \approx 0.488$$

$$P(plates, m|OK) = \frac{5}{21} \approx 0.24$$

$$P(plates, m) = P(plates, m|OK)P(OK) + \underbrace{\boxed{P(plates, m|\neg OK)}}_{8/22 \approx 0.36}\boxed{P(\neg OK)}$$

$$\frac{22}{43} \approx 0.512$$

Estimate **posterior** probability using Bayes

What is the probability of having an **OK** (or a **not OK**) shipping, if we pack **plates** in a **medium cardboard**?

$$P(OK \mid plates, m) = \frac{P(plates,m|OK)P(OK)}{P(plates,m)}$$

$$= \frac{\frac{5}{21} \cdot \frac{21}{43}}{\frac{5}{21} \cdot \frac{21}{43} + \frac{8}{22} \cdot \frac{22}{43}} = \frac{5}{13} \approx 0.38$$

object	package	OK shipping	Not OK shipping
wine	thin (t)	3	7
plates	medium (m)	5	8
plates	hard (h)	8	3
wine	medium (m)	5	4
	total:	**21**	**22**

Question: **Is it <u>OK</u> or <u>not OK</u> to ship plates in a medium package?**

$$P(OK \mid plates, m) = \frac{P(plates,m|OK)P(OK)}{P(plates,m)}$$

$$= \frac{\frac{5}{21}\cdot\frac{21}{43}}{\frac{5}{21}\cdot\frac{21}{43}+\frac{8}{22}\cdot\frac{22}{43}} = \frac{5}{13} \approx 0.38$$

object	package	OK shipping	Not OK shipping
wine	thin (t)	3	7
plates	medium (m)	5	8
plates	hard (h)	8	3
wine	medium (m)	5	4
	total:	**21**	**22**

$$P(\neg OK \mid plates, m) \approx 0.62$$

How do we classify this shipping? **Not OK**

Estimating probabilities

An example, when we do not have complete evidence

Question: **Is it OK or not OK to ship plates in a thin package?**

What is the probability of having an **OK** (or **not OK**) shipping, if we pack plates in a thin cardboard?

$$P(OK \mid plates, t) = ?$$

$$P(OK \mid plates, t) = \frac{\boxed{P(plates, t \mid OK)}\,P(OK)}{P(plates, t)}$$

We CANNOT estimate it from the data!

object	package	OK shipping	Not OK shipping
wine	thin (t)	3	7
plates	medium (m)	5	8
plates	hard (h)	8	3
wine	medium (m)	5	4
	total:	21	22

Problems with estimation of probability

Estimation of probability is a problem for high dimensional data (i.e. many features). The more features we have, the more data point we need for reliable estimations.

Curse of dimensionality
In higher-dimensional spaces, the data become too *sparse*. This creates problems as the prediction are not statistically significant (thus we need *exponentially* more data).

One **approach/solution** is to make an assumption on **statistical independence of features** (independent events): their value is not dependent on each other

Independent events

An event that has no connection to another event's chances of happening (or not happening).

The event has no effect on the probability of another event occurring.

$$P(A, B) \longrightarrow P(A)P(B)$$

Naïve Bayes: independent events

Very difficult to estimate!

Bayes rule: $$P(C|x_1, x_2, \ldots, x_n) = \frac{\boxed{P(x_1,x_2,\ldots,x_n|C)}P(C)}{P(x_1,x_2,\ldots,x_n)}$$

Assuming statistical independence:
product of probabilities

$$\boxed{P(x_1, x_2, \ldots, x_n|C) = P(x_1|C)P(x_2|C)\ldots P(x_n|C)}$$

Naïve Bayes rule:

$$P(C|x_1, x_2, \ldots, x_n) = \frac{P(x_1|C)P(x_2|C)\ldots P(x_n|C)P(C)}{P(x_1,x_2,\ldots,x_n)}$$

Naïve Bayes: advantages

Each feature distribution can be independently estimated as a 1D distribution.

No need for large data sets that scale exponentially with the number of features.

Empirical observation: **In many cases it works**.

Naïve explanation: Correct classification as long as the correct class is more probable than any other classes (hence class probabilities do not have to be estimated precisely).

...back to the shipping prediction

What is the probability of having an **OK** shipping, if we pack plates in an thin cardboard?

$$P(OK \mid plates, t) = \frac{P(plates,t|OK)P(OK)}{P(plates,t)} =$$

object	package	OK shipping	Not OK shipping
wine	thin (t)	3	7
plates	medium (m)	5	8
plates	hard (h)	8	3
wine	medium (m)	5	4
	total:	**21**	**22**

...back to the shipping prediction

What is the probability of having an **OK** shipping, if we pack plates in an thin cardboard?

$$P(OK \mid plates, t) = \frac{P(plates{,}t|OK)P(OK)}{P(plates{,}t)} = \frac{P(plates|OK)P(t|OK)P(OK)}{P(plates{,}t)}$$

object	package	OK shipping	Not OK shipping
wine	thin (t)	3	7
plates	medium (m)	5	8
plates	hard (h)	8	3
wine	medium (m)	5	4
	total:	**21**	**22**

...back to the shipping prediction

What is the probability of having an **OK** shipping, if we pack plates in an thin cardboard?

$$P(OK \mid plates, t) = \frac{P(plates,t|OK)P(OK)}{P(plates,t)} = \frac{P(plates|OK)P(t|OK)P(OK)}{P(plates,t)}$$

object	package	OK shipping	Not OK shipping
wine	thin (t)	3	7
plates	medium (m)	5	8
plates	hard (h)	8	3
wine	medium (m)	5	4
	total:	**21**	**22**

$$P(plates|OK) = ?$$

$$P(t|OK) = ?$$

$$P(plates, t) = ?$$

...back to the shipping prediction

What is the probability of having an **OK** shipping, if we pack plates in an thin cardboard?

$$P(OK \mid plates, t) = \frac{P(plates,t|OK)P(OK)}{P(plates,t)} = \frac{P(plates|OK)P(t|OK)P(OK)}{P(plates,t)}$$

object	package	OK shipping	Not OK shipping
wine	thin (t)	3	7
plates	medium (m)	5	8
plates	hard (h)	8	3
wine	medium (m)	5	4
	total:	**21**	**22**

$$P(plates|OK) = \frac{13}{21}$$

$$P(t|OK) = ?$$

$$P(plates, t) = ?$$

...back to the shipping prediction

What is the probability of having an **OK** shipping, if we pack <u>plates</u> in an <u>thin</u> cardboard?

$$P(OK \mid plates, t) = \frac{P(plates,t|OK)P(OK)}{P(plates,t)} = \frac{P(plates|OK)P(t|OK)P(OK)}{P(plates,t)}$$

object	package	OK shipping	Not OK shipping
wine	thin (t)	$\boxed{3}$	7
plates	medium (m)	5	8
plates	hard (h)	8	3
wine	medium (m)	5	4
	total:	$\boxed{\mathbf{21}}$	**22**

$$P(plates|OK) = \frac{13}{21}$$

$$P(t|OK) = \frac{3}{21}$$

$$P(plates, t) = ?$$

...back to the shipping prediction

What is the probability of having an **OK** shipping, if we pack plates in an thin cardboard?

$$P(OK \mid plates, t) = \frac{P(plates,t|OK)P(OK)}{P(plates,t)} = \frac{P(plates|OK)P(t|OK)P(OK)}{P(plates,t)}$$

$$P(plates, t) = P(plates|OK)P(t|OK)P(OK) + P(plates|\neg OK)P(t|\neg OK)P(\neg OK)$$

$$= \frac{13}{21} \cdot \frac{3}{21} \cdot \frac{21}{43} + \frac{11}{22} \cdot \frac{7}{22} \cdot \frac{22}{43}$$

...back to the shipping prediction

What is the probability of having an **OK** shipping, if we pack plates in an thin cardboard?

$$P(OK \mid plates, t) = \frac{P(plates, t|OK)P(OK)}{P(plates, t)} = \frac{P(plates|OK)P(t|OK)P(OK)}{P(plates, t)}$$

$$= \frac{\frac{13}{21} \cdot \frac{3}{21} \cdot \frac{21}{43}}{\frac{13}{21} \cdot \frac{3}{21} \cdot \frac{21}{43} + \frac{11}{22} \cdot \frac{7}{22} \cdot \frac{22}{43}} \approx 0.34$$

...back to the shipping prediction

What is the probability of having an **OK** shipping, if we pack plates in an thin cardboard?

$$P(OK \mid plates, t) = \frac{P(plates,t|OK)P(OK)}{P(plates,t)} = \frac{P(plates|OK)P(t|OK)P(OK)}{P(plates,t)}$$

$$= \frac{\frac{13}{21}\cdot\frac{3}{21}\cdot\frac{21}{43}}{\frac{13}{21}\cdot\frac{3}{21}\cdot\frac{21}{43}+\frac{11}{22}\cdot\frac{7}{22}\cdot\frac{22}{43}} \approx 0.34$$

Compute similarly for the event 'shipping class **not OK**':

$$P(\neg OK|plates, t) \approx 0.66$$

...back to the shipping prediction

$$P(OK \mid plates, t) = \frac{P(plates, t|OK)P(OK)}{P(plates, t)} = \frac{\boxed{P(plates|OK)P(t|OK)P(OK)}}{P(plates, t)}$$

$$P(\neg OK \mid plates, t) = \frac{P(plates, t|\neg OK)P(\neg OK)}{P(plates, t)} = \frac{\boxed{P(plates|\neg OK)P(t|\neg OK)P(\neg OK)}}{P(plates, t)}$$

$$P(OK \mid plates, t) > P(\neg OK \mid plates, t) \longrightarrow$$ **Choose OK**

$$P(OK \mid plates, t) < P(\neg OK \mid plates, t) \longrightarrow$$ **Choose Not OK**

Summary

- To interact, machine have to sense and understand the world
- Machines learn a representation of the world from examples
- We can model world events using probability (Bayes theory)
- We can make predictions also when we do not have complete training data
    - We make assumptions on the data (statistical independence of features -> Naïve Bayes)

Assignments

This afternoon, focus on the **preparative assignment 7.1**

Start working on the **pearl assignment 7.5**