Introduction to intelligent systems

Algorithmic fairness

Mikkel N. Schmidt

Technical University of Denmark,
DTU Compute, Department of Applied Mathematics and Computer Science.

Overview

1. Fairness
2. AI ethics
3. Recap
   - Statistics
   - Machine learning
   - Algorithms
   - Optimization
   - Data
4. Written exam
5. Project work
   - Activities in the project period

Feedback group

- Casper Andresen
- Jasmin Lundager Aasbjerg Petersen
- Gabriel Sejr Hornstrup
- Sofus Alexander Kjelgaard Carstens

Learning objectives

- II Fairness criteria: Demographic parity, equalized odds, equal opportunity.
- I Ethical challenges and dilemmas in AI

- I Understand the concepts and definitions, and know their application. Reason about the concepts in the context of an example. Use correct technical terminology.
- II As above plus: Read, manipulate, and work with technical definitions and expressions (mathematical and Python code). Carry out practical computations. Interpret and evaluate results.

Fairness

Setting and notation

We restrict ourselves to a binary decision process setting, with two protected groups

$G \in \{a, b\}$	Protected group (such as race, gender, religion etc.)
X	Attributes used in decision making
$\hat{S} = f(X) \in \mathbb{R}$	Score function
$\hat{Y} \in \{0, 1\}$	Binary decision found by thresholding $\hat{S}$
$Y \in \{0, 1\}$	Correct decision (might not be available)

Binary decision making

- Score function

$$\hat{S} = f(X)$$

E.g. determined by machine learning to predict a measure of success.

- Binary decision

$$\hat{Y} = \begin{cases} 1 & \text{if } \hat{S} > \tau, \\ 0 & \text{otherwise.} \end{cases}$$

Decision matrix

Decision matrix

	$G=a$		$G=b$	
	$\hat{Y}=0$	$\hat{Y}=1$	$\hat{Y}=0$	$\hat{Y}=1$
$Y=0$	a_{00}	a_{01}	b_{00}	b_{01}
$Y=1$	a_{10}	a_{11}	b_{10}	b_{11}

- Our decision $\hat{Y}$ is known.
- Outcome Y sometimes only known for $\hat{Y}=1$.
- Outcome Y for $\hat{Y}=0$ can be a counterfactual.

Demographic parity

Decide $\hat{Y}=1$ and $\hat{Y}=0$ in the same fraction of cases in each protected group

$$P(\hat{Y}=1|G=a) = P(\hat{Y}=1|G=b)$$

	$G=a$		$G=b$	
	$\hat{Y}=0$	$\hat{Y}=1$	$\hat{Y}=0$	$\hat{Y}=1$
$Y=0$	a_{00}	a_{01}	b_{00}	b_{01}
$Y=1$	a_{10}	a_{11}	b_{10}	b_{11}

Selection of competitors for the math competition

- A school participates in a math competition and can run with 12 students.
- The school has students from two protected groups
  - a: (200 students) and
  - b: (100 students).

We have access to the students' assessment marks $\hat{S}$, which we believe is a reasonable predictor of success.
How can we select students in a fair manner according to the demographic parity criterion?

Selection of competitors for the math competition

- A school participates in a math competition and can run with 12 students.
- The school has students from two protected groups
  - a: (200 students) and
  - b: (100 students).

We have access to the students' assessment marks $\hat{S}$, which we believe is a reasonable predictor of success.
How can we select students in a fair manner according to the demographic parity criterion?

Solution
We can select $\frac{200}{200+100} \cdot 12 = 8$ from group a and the remaining 4 from group b, ranked by their assessment marks.

Equalized odds

Probability of correct decision equal in the protected groups

$$P(\hat{Y}=0|Y=0,G=a)=P(\hat{Y}=0|Y=0,G=b)$$
$$P(\hat{Y}=1|Y=1,G=a)=P(\hat{Y}=1|Y=1,G=b)$$

	$G=a$		$G=b$	
	$\hat{Y}=0$	$\hat{Y}=1$	$\hat{Y}=0$	$\hat{Y}=1$
$Y=0$	a_{00}	a_{01}	b_{00}	b_{01}
$Y=1$	a_{10}	a_{11}	b_{10}	b_{11}

Selection of competitors for the math competition II

After the math competition, the math problems are released and all students have a go. It turns out that students in group b were a lot better on average, and that the assessment marks we not a good predictor of success.

	$G = a$		$G = b$	
	$\hat{Y} = 0$	$\hat{Y} = 1$	$\hat{Y} = 0$	$\hat{Y} = 1$
$Y = 0$	96	4	24	1
$Y = 1$	96	4	72	3

With this new data in mind, was the school's selection fair according to the equalized odds criterion?

Selection of competitors for the math competition II

After the math competition, the math problems are released and all students have a go. It turns out that students in group b were a lot better on average, and that the assessment marks we not a good predictor of success.

	$G=a$		$G=b$	
	$\hat{Y}=0$	$\hat{Y}=1$	$\hat{Y}=0$	$\hat{Y}=1$
$Y=0$	96	4	24	1
$Y=1$	96	4	72	3

With this new data in mind, was the school's selection fair according to the equalized odds criterion?

Solution

$$P(\hat{Y}=0|Y=0,G=a)=P(\hat{Y}=0|Y=0,G=b)$$
$$\frac{96}{96+4}=\frac{24}{24+5}=96\%$$
$$P(\hat{Y}=1|Y=1,G=a)=P(\hat{Y}=1|Y=1,G=b)$$
$$\frac{4}{96+4}=\frac{3}{72+3}=4\%$$

Yes, this if fair according to equalized odds.

Equalized opportunity

Probability that decision is correct among positive decisions equal in the protected groups

$$P(Y=1 \mid \hat{Y}=1, G=a) = P(Y=1 \mid \hat{Y}=1, G=b)$$

	$G=a$		$G=b$	
	$\hat{Y}=0$	$\hat{Y}=1$	$\hat{Y}=0$	$\hat{Y}=1$
$Y=0$	a_{00}	a_{01}	b_{00}	b_{01}
$Y=1$	a_{10}	a_{11}	b_{10}	b_{11}

Selection of competitors for the math competition III

	$G = a$		$G = b$	
	$\hat{Y} = 0$	$\hat{Y} = 1$	$\hat{Y} = 0$	$\hat{Y} = 1$
$Y = 0$	96	4	24	1
$Y = 1$	96	4	72	3

Was the school's selection fair according to the equalized opportunity criterion?

Selection of competitors for the math competition III

	$G = a$		$G = b$	
	$\hat{Y} = 0$	$\hat{Y} = 1$	$\hat{Y} = 0$	$\hat{Y} = 1$
$Y = 0$	96	4	24	1
$Y = 1$	96	4	72	3

Was the school's selection fair according to the equalized opportunity criterion?

Solution

$$P(Y = 1 | \hat{Y} = 1, G = a) = P(Y = 1 | \hat{Y} = 1, G = b)$$

$$\frac{4}{4+4} = 50\% = \frac{3}{1+3} = 75\% \text{ False}$$

No, this does not appear to be fair according to equalized opportunity. Caveat: The sample size is very small, so the difference is not statistically significant.

AI ethics

Ethical challenges in AI

- Increased reliance on AI
- Explainable AI / right to explanation
- Bias and fairness in AI systems
- Behaviour manipulation
- Human-robot interaction
- Autonomous systems
- AI faking technologies
- Automation and employment
- Privacy and surveillance
- Strong AI, super intelligence, robot rights

Increased reliance on AI systems

Boeing anti-stall patch MCAS (Maneuvering Characteristics Augmentation System).

- As an automated corrective measure, the MCAS was given full authority to bring the aircraft nose down, and could not be overridden by pilot resistance against the control wheel.
- Pilots were unaware of the existence of MCAS due to its omission from the crew manual and no coverage in training.[1]
- In October 2018 and March 2019, Boeing 737 MAX passenger jets crashed just after takeoff with nearly 350 killed.
- All 737 MAX planes were afterwards grounded worldwide.

[1] Wikipedia, https://en.wikipedia.org/wiki/Maneuvering_Characteristics_Augmentation_System

Increased reliance on AI systems

Boeing anti-stall patch MCAS (Maneuvering Characteristics Augmentation System).

- As an automated corrective measure, the MCAS was given full authority to bring the aircraft nose down, and could not be overridden by pilot resistance against the control wheel.
- Pilots were unaware of the existence of MCAS due to its omission from the crew manual and no coverage in training.[1]
- In October 2018 and March 2019, Boeing 737 MAX passenger jets crashed just after takeoff with nearly 350 killed.
- All 737 MAX planes were afterwards grounded worldwide.

> *"[MCAS] was designed per our standards, certified per our standards, and we're confident in that process. So, it operated according to those design and certification standards. So, we haven't seen a technical slip or gap in terms of the fundamental design and certification of the approach." (Boeing CEO Dennis Muilenburg 2019)*

[1] Wikipedia, https://en.wikipedia.org/wiki/Maneuvering_Characteristics_Augmentation_System

Questions

- In what daily situations do you rely on automation and AI?
- Is it a problem if humans rely on computer decisions without understanding how they operate or draw conclusions?
- Should humans be made aware of all AI systems they come in contact with? And should they understand how they work?

Dilemma

You use a chatbot to make phone calls to order a service on your behalf. The service provider never finds out that they talked to a bot. Is that okay?

- **Yes** As long as everything works for everybody
- **No** It has to be clear if you talk to a human or a robot

Explainable AI

- At ML Symposium, NIPS 2017, Facebook chief AI Scientist Yann LeCun suggests that a models reasoning can be inferred from observing how it acts, such that:
  - Rigorous testing is enough to provide an explanation.
- Cassie Kozyrkov (Google Intelligence Engineer) asks the question: "Imagine choosing between two spaceships.
  - Spaceship 1 comes with exact equations explaining how it works, but has never been flown.
  - How Spaceship 2 flies is a mystery, but it has undergone extensive testing, with years of successful flights like the one you're going on.

  Which spaceship would you choose?"

Questions

- If the model is predictable and thoroughly tested, do you think there is still a need for an explanation?
- What in cases were obtaining testdata for all cases is infeasible or costly (healthcare, self-driving cars)?
- Do you hold other humans to the same standards?

Dilemma

Use of AI can lead to faster decisions within the public sector, such as building permissions and early retirement benefits.

Good idea It makes processing times faster.

Bad idea Cases should be processed by humans.

Right to explanation

Generally in the EU, consumers have a legal *right to explanation* if a decision is based solely on automated processing (including profiling). Profiling can be the basis for decision making in certain situations (such as fraud and tax-evasion monitoring), or when the data subject has given his or her explicit consent.

- Do you think this is a fair principle?
- Should "right to explanation" limit what AI models we can use for decision making?
- Should this also be the case when humans make the decisions?

Bias in AI systems

- In 2014 Amazon developed a machine learning system to review job applicants' resumes to automatically extract the top talents.
- The model was trained to score applicants based on patterns observed in resumes the company had received for over a decade.
- Most of the training data were applications from men, reflecting the demographics of the tech industry.

[2] Reuters, Oct 11, 2018, https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G

Bias in AI systems

- In 2014 Amazon developed a machine learning system to review job applicants' resumes to automatically extract the top talents.
- The model was trained to score applicants based on patterns observed in resumes the company had received for over a decade.
- Most of the training data were applications from men, reflecting the demographics of the tech industry.

"In effect, Amazon's system taught itself that male candidates were preferable. It penalized resumes that included the word women's, as in women's chess club captain. And it downgraded graduates of two all-women's colleges [...] Amazon edited the programs to make them neutral to these particular terms. But that was no guarantee that the machines would not devise other ways of sorting candidates that could prove discriminatory, the people said."[2]

[2] Reuters, Oct 11, 2018, https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G

Questions

AI systems learns from training data, which can reflect biased human decisions or historical constructs, even when sensitive variables are removed.

- How should we address this?
- Is it possible to agree on when a decision is fair?

Behaviour manipulation

- On March 23, 2016, Microsoft launched *Tay*, an artificial intelligence chatbot on Twitter, developed to conduct research on conversational understanding.
- Tay was designed to mimic the language of a young adult learning by interacting with human Twitter users.
- Users could follow and interact with the bot on Twitter and it wood tweet back, learning conversation from other users' posts.
- Just 16 hours after its launch it was shut down

Behaviour manipulation

- On March 23, 2016, Microsoft launched *Tay*, an artificial intelligence chatbot on Twitter, developed to conduct research on conversational understanding.
- Tay was designed to mimic the language of a young adult learning by interacting with human Twitter users.
- Users could follow and interact with the bot on Twitter and it wood tweet back, learning conversation from other users' posts.
- Just 16 hours after its launch it was shut down

  "As many of you know by now, on Wednesday we launched a chatbot called Tay. We are deeply sorry for the unintended offensive and hurtful tweets from Tay, which do not represent who we are or what we stand for, nor how we designed Tay. Tay is now offline and we'll look to bring Tay back only when we are confident we can better anticipate malicious intent that conflicts with our principles and values." (Mar 25, 2016, Peter Lee, Corporate Vice President, Microsoft Healthcare)

Questions

- Who is responsible for Tay's behaviour?

Questions

- Who is responsible for Tay's behaviour?

Just as the AI's behaviour can be manipulated, so can humans. Manipulating online behaviour is a core internet business model, which includes exploitation of behaviour, addiction and deception.

- Sale of addictive goods and services (alcohol, tobacco, gambling) are highly regulated, should this be the same for technology that exploit or manipulate behaviour?
- The Hippocratic oath historically taken by physicians states "I will abstain from all intentional wrong-doing and harm". Do we have similar ethical obligations as AI engineers?

Human-Robot interactions

- "A California hospital delivered end-of-life news to a 78-year-old patient via a robotic machine this week, prompting the man's family to go public with their frustration."[3]

[3] Headline, USAtoday 2018

[4] Aalborg University, https://www.kommunikation.aau.dk/forskning/vidensgrupper/e-learning-lab/nyhedsliste/nyhed/demente-liver-op-i-interaktion-med-menneskelignende-robotter.cid365882

Human-Robot interactions

- "A California hospital delivered end-of-life news to a 78-year-old patient via a robotic machine this week, prompting the man's family to go public with their frustration."[3]
- People with severe dementia benefit from contact with social robots: It makes them come alive and socially active. But there is a risk that the robots may be used to passify troublesome patients.[4]

[3] Headline, USAtoday 2018

[4] Aalborg University, https://www.kommunikation.aau.dk/forskning/vidensgrupper/e-learning-lab/nyhedsliste/nyhed/demente-liver-op-i-interaktion-med-menneskelignende-robotter.cid365882

Dilemma

Is it okay that a nursing home uses robot pets to interact with residents with dementia, who do not know it is a robot?

Yes It brings joy and increases life quality

No Residents need to know if they interact with robots

Recap

Overview

Statistics

- Population and sample
- Standard error and confidence intervals
- Sample size calculation
- Correlation and causality
- Experimental design
- Fairness criteria

ML

- Unsupervised, supervised, and reinforcement learning
- Training and test error
- Feature transformation: Scaling and basis change

Algorithms

- Algorithmic complexity
- K-means clustering
- Least squares regression
- Neural networks
- TF-IDF / Okapi BM25
- Value iteration / q-learning

Optimiz.

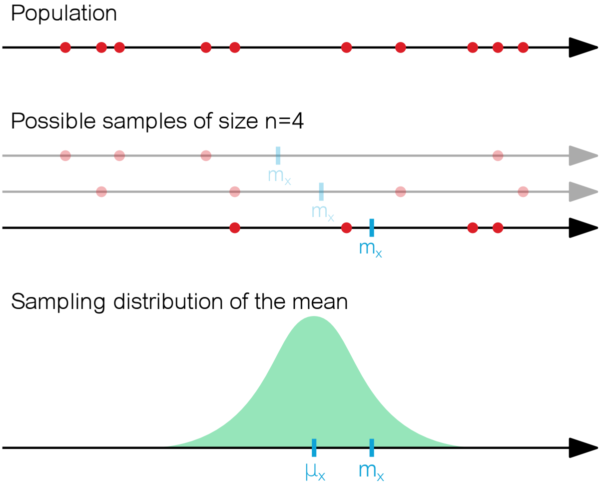
- Cost function and parameter estimation
- Gradient descent
- Automatic differentiation

Data

- Image data: Color spaces
- Audio data: Spectrogram
- Text data: Bag of words

Recap: Statistics

Population and sample



Standard error and confidence intervals

Confidence interval

point estimate $\pm$ margin of error

$$\bar{x} \pm \underbrace{z_{\alpha/2}}_{\text{critical value}} \cdot \underbrace{\sqrt{\frac{\sigma^2}{n}}}_{\text{standard error}}$$

We want to estimate the population mean μ

- Sample n observations
- Compute the sample mean $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
- Choose confidence level, e.g. $1 - \alpha = 95\%$, and look up critical value
- Compute standard error and multiply by critical value

Sample size calculation

- The equation for the confidence interval can be solved for the sample size
- This gives a formula for the required sample size to give a desired margin of error

Sample size for proportion

$$n = z_{\alpha/2}^2 \frac{p(1-p)}{E^2}$$

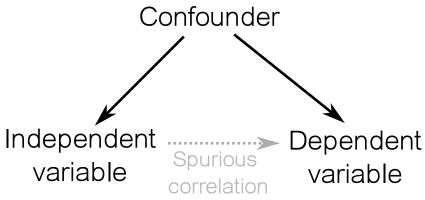
E: Desired margin of error
α: Significance level
p: True proportion

Correlation and causality

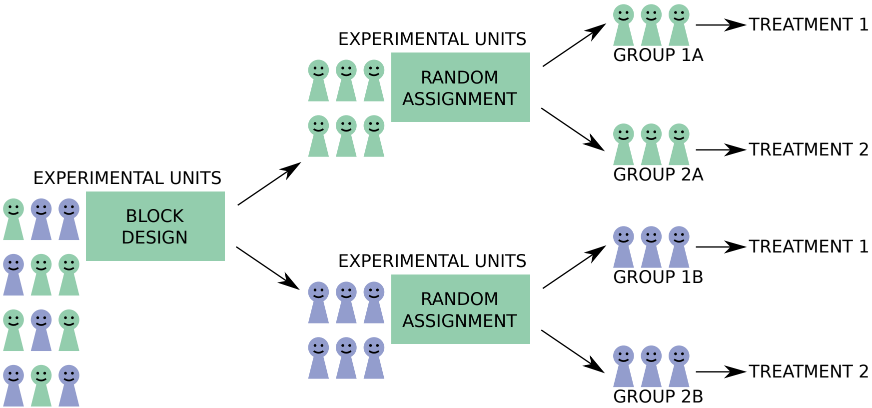
Pearson correlation coefficient

$$\rho_{X,Y} = \frac{\mathrm{cov}(X, Y)}{\sigma_X \cdot \sigma_Y}$$

Causality



Experimental design



- Blocking eliminates one confounding variable
- Other confounders will be randomly distributed between treatment groups

Recap: Machine learning

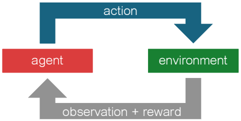
Unsupervised, supervised, and reinforcement learning

Categorization of learning problems

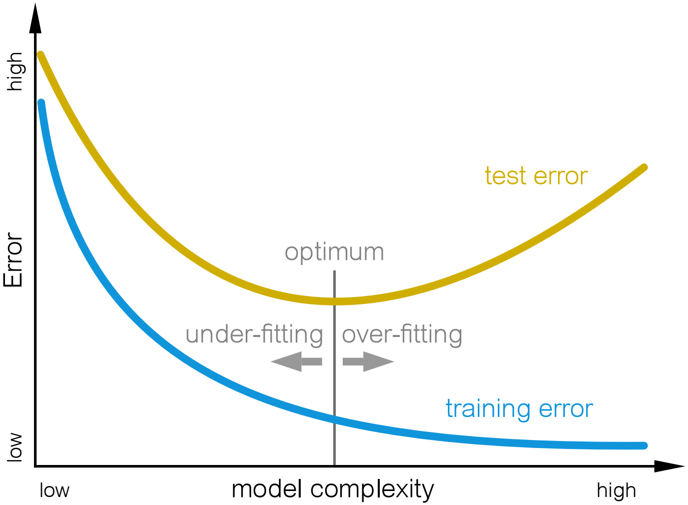
Unsupervised Learn function that describes the structure in data

Supervised Learn function that maps input to output to optimize cost

Reinforcement Learn a function (policy) that maps inputs to actions to optimize cumulative reward



Training and test error



Feature transformation: Scaling and basis change

Min-max normalization

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

Standardization

$$x' = \frac{x - \bar{x}}{\sigma_x}$$

- $\bar{x}$ Mean
- σ_x standard deviation

Basis change

$$\boldsymbol{y} = \underbrace{\boldsymbol{T}^{-1}\boldsymbol{x}}_{\text{matrix multiplication}}$$

- $\boldsymbol{x}$ Vector in the original coordinate system
- $\boldsymbol{T}$ Matrix where each column is a basis vector of the new coordinate system expressed in the old coordinate system
- $\boldsymbol{y}$ Vector expressed in the new coordinate system

Recap: Algorithms

Algorithmic complexity

- Classify algorithms according to their performance
- Time function $T(n)$ measures runtime
- Big-O notation expresses *runtime complexity*
- Considers only the highest order term of $T(n)$
- Upper bound on growth rate

Formal definition

$T(n) \in O(f(n))$ iff there exists a constant c such that $T(n) < c\,f(n)$ for all $n > n_0$
We say $f(n)$ is an asymptotic upper bound for $T(n)$

Sipser, Introduction to the theory of computation, 2006

K-means clustering

Objective

$$\underbrace{\min_{\{\mu_1,\ldots,\mu_K\}}}_{\text{Cluster means}} \underbrace{\min_{\{c_1,\ldots,c_N\}}}_{\text{Cluster assignments}} \underbrace{\sum_{n=1}^{N} \|x_n - \mu_{c_n}\|^2}_{\text{Squared distance to cluster mean}}$$

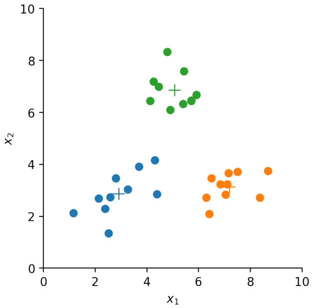
Algorithm

1. Fix cluster means, optimize cluster assignment

$$\min_{\{c_1,\ldots,c_N\}} \sum_{n=1}^{N} \|x_n - \mu_{c_n}\|^2$$

2. Fix cluster assignments, optimize cluster means

$$\min_{\{\mu_1,\ldots,\mu_K\}} \sum_{n=1}^{N} \|x_n - \mu_{c_n}\|^2$$

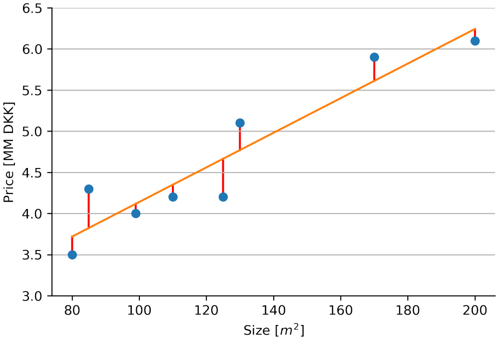


Least squares regression

- Regression line: $f(x) = ax + b$
- Error: Squared distance between data and regression line

$$E = \sum_{n=1}^{N} (y_n - f(x_n))^2$$

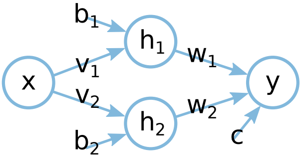
- Find values of a and b to minimize E



Neural networks

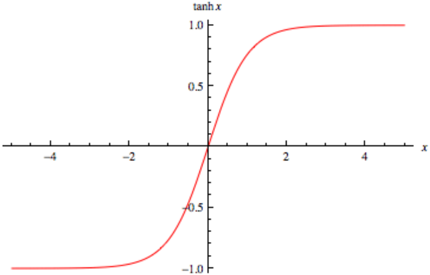
Cost function

$$E = \sum_{n=1}^{N} \big(y(n) - \hat{y}(n)\big)^2$$



Network structure

$$\begin{aligned} \hat{y}(n) &= w_1 h_1(n) + w_2 h_2(n) + c \\ h_1(n) &= \tanh\left(v_1 x(n) + b_1\right) \\ h_2(n) &= \tanh\left(v_2 x(n) + b_2\right) \end{aligned}$$



Model parameters

$$c,\ w_1,\ w_2,\ v_1,\ v_2,\ b_1,\ b_2$$

Okapi BM25

$$\mathrm{BM25}(d, q) = \sum_{t \in q} \frac{n_{t,d} \cdot (k_1 + 1)}{n_{t,d} + k_1 \cdot (1 - b + b \cdot \frac{n_d}{\mathrm{avgdl}})} \cdot \log\left(\frac{N - n_t + 0.5}{n_t + 0.5}\right)$$

- $n_{t,d}$ Number of occurences of term t in document d
- n_d Number of terms in document d
- n_t Number of documents with term t
- N Total number of documents
- avgdl Average document length $\frac{1}{N} \sum_d n_d$
- b Parameter ($b \in [0, 1]$, default $b = 0.75$)
- k_1 Parameter ($k_1 > 0$, default $k_1 = 1.2$)

Britta Weber, "BM25 demystified", https://www.youtube.com/watch?v=v3Ko0CwgTZ0

Stephen Robertson and Hugo Zaragoza, "The probabilistic relevance framework: BM25 and beyond"

Wikipedia, https://en.wikipedia.org/wiki/Okapi_BM25

Value iteration and Q-learning

Value iteration

- Loop through all states and update according to

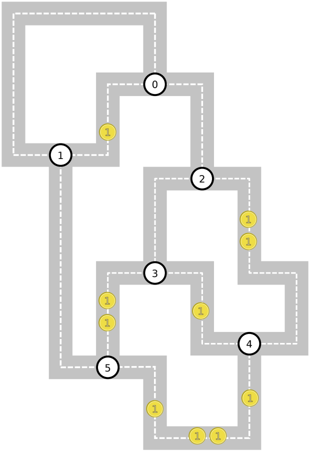
$$v(s) = \max_a \left(r(s,a) + \gamma v(s')\right)$$

- Repeat until convergence

Q-learning

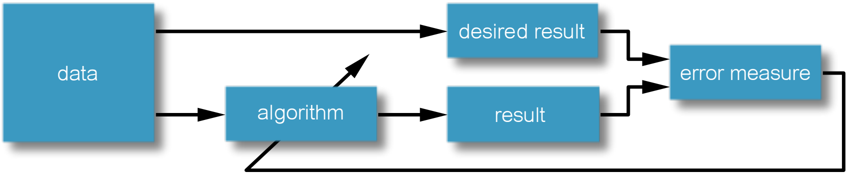
- Explore the environment according to some policy that ensures visiting all state-action pair
- At each step, update the quality function according to

$$q(s,a) = r(s,a) + \gamma \max_{a'} q(s', a')$$

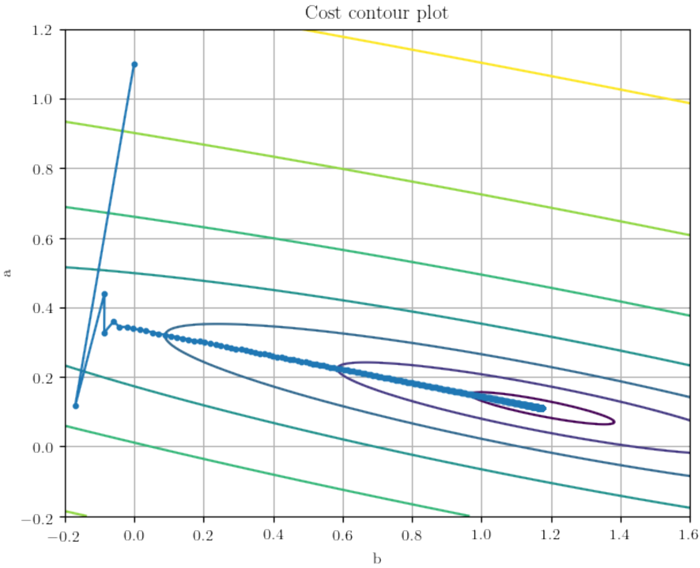


Recap: Optimization

Cost function and parameter estimation



Gradient descent

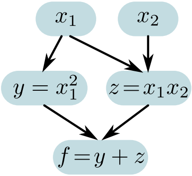


Automatic differentiation

Function and derivatives

$$f(x_1, x_2) = x_1^2 + x_1 \cdot x_2$$

$$\frac{\partial f}{\partial x_1} = 2x_1 + x_2, \quad \frac{\partial f}{\partial x_2} = x_1, \quad \nabla f(3,4) = \begin{bmatrix} 10 \\ 3 \end{bmatrix}$$



Evaluate $f(3,4)$

$$\begin{aligned} x_1 &= 3 \\ x_2 &= 4 \\ y &= x_1^2 \\ z &= x_1 x_2 \\ f &= y + z \end{aligned}$$

Evaluate $\nabla f(3,4)$

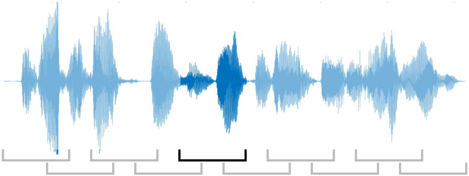
$$\begin{aligned} \bar{f} &= 1 \\ \bar{y} &= \bar{y} + \bar{f}\frac{\partial f}{\partial y} = \bar{y} + \bar{f} \cdot 1 = 0 + 1 \cdot 1 = 1 \\ \bar{z} &= \bar{z} + \bar{f}\frac{\partial f}{\partial z} = \bar{z} + \bar{f} \cdot 1 = 0 + 1 \cdot 1 = 1 \\ \bar{x}_1 &= \bar{x}_1 + \bar{y}\frac{\partial y}{\partial x_1} = \bar{x}_1 + \bar{y} \cdot 2 \cdot x_1 = 0 + 1 \cdot 2 \cdot 3 = 6 \\ \bar{x}_1 &= \bar{x}_1 + \bar{z}\frac{\partial z}{\partial x_1} = x_1 + \bar{z} \cdot x_2 = 6 + 1 \cdot 4 = 10 \\ \bar{x}_2 &= \bar{x}_2 + \bar{z}\frac{\partial z}{\partial x_2} = x_2 + \bar{z} \cdot x_1 = 0 + 1 \cdot 3 = 3 \end{aligned}$$

Recap: Data

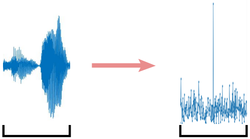
Image data: Color spaces

Audio data: Spectrogram

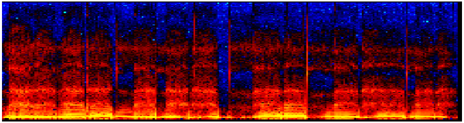
Split the signal
into blocks



For each block,
compute the spectrum



Gather spectra as columns
in matrix and plot heat map



Text data: Bag of words

	Doc. 1	Doc. 2
african	1	0
although	0	1
and	1	0
are	1	0
bear	0	1
black	1	0
by	1	0
close	0	1
coat	1	0
distinct	1	0
equid	1	0
famili	1	0
giraff	0	1
hors	1	0
is	0	1
it	0	1
mark	0	1
most	0	1
of	1	1
okapi	0	1
relat	0	1
reminisc	0	1
sever	1	0
speci	1	0
stripe	1	1
the	0	2
their	1	0
to	0	1
unit	1	0
white	1	0
zebra	1	1

Sentences

1. Zebras are several species of African equids (horse family) united by their distinctive black and white striped coats.
2. Although the okapi bears striped markings reminiscent of zebras it is most closely related to the giraffe.

- A bag-of-words sentence/document can be seen as a point in a high-dimensional vector space

Written exam

Individual written examination

Written exam in December

- Topics covered in lecture slides and notes (see “list of topics”)
- 2 hours
- 5 exercises with 4 questions each
- Hand in digitally (pdf file)

Final grade

- Written exam (weight 40%)
- Individualized group report (weight 60%)

List of topics

Show list of topics

Project work

Project work

- Groups of 3 students is strongly preferred
- Design and conduct an experiment
- Write and hand in group report.
  - 5-pages.
  - Individualized (who did what?)

Week 1

Come up with idea The first step is to come up with an idea for your experiment—something you would like to study. You can formulate this as a research question, and state your expectations and tentative explanation as a hypothesis.

Design experiment Next, you must design your experiment. Agree on a protocol to follow and write it down. Ideally, the experiment should be designed so carefully that others would be able to reproduce the experiment and get the same result except for statistical variation.

Decide on your approach and start building At this time, you should also agree on which method to use to analyze the data, and consider how large a sample size you need to support your claims. You should think about what code you need to write, or existing toolboxes you need to familiarize yourself with, to be able to conduct your experiment. Start prototyping your experimental setup.

Deliverables

- Description of experiment idea
- Experimental plan

Week 2

Build your experimental platform At this time, you should put together the computer code etc. you need for your experiment. Test that everything works as required and finalize your setup. Perhaps it could be a good idea to run a small pilot study to make sure everything is in place.

Carry out your experiment Now, you are ready to carry out your experiment. Follow your protocol carefully and record the results. If something goes wrong or you realize there is an issue with your experimental design, you might need to go back to the previous step and modify your design.

Start writing At this time, you should also start to draft your report.

Deliverables

- Experimental data
- Report draft

Week 3

Document and communicate your experiment Finally, you must write up a report, following the template we have given you. The report must be approximately 5 pages long, and you must hand it in as a pdf file.
The report must clearly describe the experimental design and protocol, include visualization and summaries of the data gathered in the experiment, present the reults of the experiment, and include a discussion where you comment on perspectives, ethics etc.

Deliverables

- Final report

Experiment

- The experiment must be centered around a problem where artificial intelligence or machine learning is relevant
- The experiment can be based on the methods and computer code discussed in the course, including
  - Image classification
  - Symbolic AI
  - Linear / neural network regression
  - K-means clustering
  - Text search using Okapi
  - Audio classification
  - Tabular value iteration / q-learning
- You must gather your own experimental data

Report

- Approximately 5 pages
- Follow the IMRaD format
- Include a short abstract
- Include appropriate visualization of the dataset
- Include statistical considerations regarding sample size
- Back up claims regarding recognition up by statistics
- Include your own code as an appendix
- Write using LaTeX

Assessment criteria

In addition to the requirements above, the degree of fulfillment of following objectives will be taken into account in the assessment of the project:

- Description of key components of intelligent systems: Sensing and active data collection, machine learning, evaluations and communication
- Application of AI tools to data (such as image, audio, text and games)
- Discussion and analysis of performance
- Application of visualization techniques for evaluation of performance and basic debugging
- Application of scientific Python programming tools
- Discusssion of the role of AI tools in the chosen application domain
- Discussion safety and ethical challenges in AI, biases and stereotypes, privacy and societal impact

Project work: Activities in the project period

Project period

Assistance	Our teaching assistants will be available many days throughout the period. They can help with almost everything, and you can book individual consultations with them.
Consultations	Groups can book individual consultations with the teacher in 20 minute time slots.
Feedback sessions	Two times during the period we will meet 4–5 groups together with the teacher to present project status and give each other feedback.
Recap lectures	Depending on your demand, we will give recap lectures in any topic from the course or beyond.

Project idea brain storm

Discuss in groups

Phase I: Generate ideas

- Present and write down very briefly all your ideas, and come up with more
- Critisism is not allowed

Phase II: Review, select, combine, and improve

- Lead a discussion to identify the most interesting ideas and try to combine or improve the ideas
- Agree on two ideas
    1. The most impactful and world changing idea.
    2. The most absurdly hillarious idea.

Be prepared to present to the class.

Report your project groups to us

- When you decide on your idea and form a group, please self-sign-up you group on DTU Learn
- Please use the discussion forum “Group formation and idea exchange” on DTU Learn
  - If you have an idea, but miss a few group members
  - If you lack ideas and group members
- It would be great if groups are formed before we start in January.

Project topic ideas

Visual learning
- Image recognition in variable lightning conditions / with obstruction
- Recognition of rotated objects
- Face recognition / facial expression recognition

Audio learning
- Classify presence or absence of music
- Noise level vs subjective noise level
- Tone / instrument classification

Text modeling
- Okapi search performance evaluation (e.g. in parilament documents or Wikipedia)
- Sentiment scoring compared with subjective evaluation (e.g parliament opening speeches or RSS news feed),

Other modalities
- EEG analysis
- Cardiac monitoring

Project examples

Emotional image analysis	Image classification with convolutional neural network and transfer learning
Generating hand-drawn circles	Image generation using generative adversarial network
Arousal in danish news media	Sentiment analysis of news articles using the AFINN lexicon
Learn to play "Game 2048"	Deep reinforcement learning using policy gradient methods
Super-resolution imaging	Image up-scaling using generative adversarial network
Decoding mental states in EEG	Binary classification using support vector machine
Recognition of hand-written digits	Image classification using convolutional neural network
Face recognition	How many training images are needed for face recognition using convolutional neural network
Learn to play "Snake"	Reinforcement learning using tabular Q-learning
Colorizing black/white images	Comparison of convolutional neural network and U-net on image colorization task
Counting objects in images	Exploration of multi-task convolutional neural network for counting objects in images
Most readable background color	Comparison of nearest neighbor, logistic regression, and neural network for classifying text readability

Thank you

See you in January