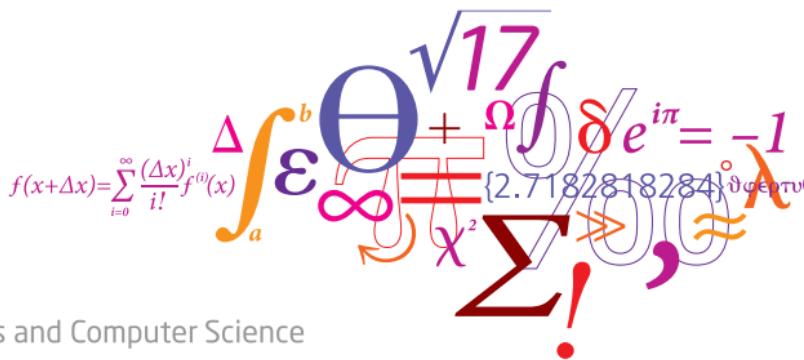


02450: Introduction to Machine Learning and Data Mining

Measures of similarity, summary statistics and probabilities

Bjørn Sand Jensen

DTU Compute, Technical University of Denmark (DTU)



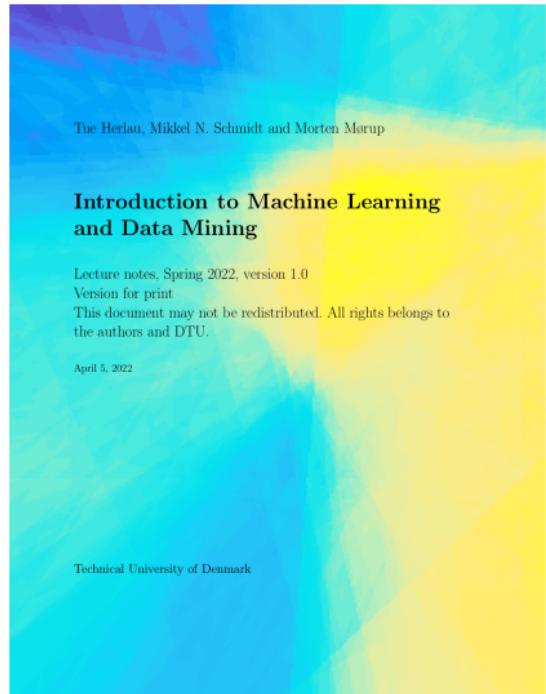
Today

Feedback Groups of the day:

Bassel Mohammad Beiroumi, Lau Bengaard, Simon Vonger Berg, Lin Bigom-Eriksen, Heanan Bird, Jonas Røkke Bjørnson, Mathilde Rosholm Block, Anna Keisha Mercy Kyei Boateng, Julie Kofod Boel, Simon Bom, Monika Karolina Borkowska, Andrea Bosisio, Laurits Bøtkjær, Emma Haley Worth Brandt, Anders Brandt-Jørgensen, Oliver Brisson Bredel, Adam Greve Bregenhof, Magnus Eriksen Brink, Oliver Birkmose Broager, Signe Wobeser Brødsgaard, Jon Brygge, Claudia Buck, Steffen Buers, Lucas Loua Buhelt, Bogdan Capsa, Alba Castrillo Perote, Juan Cervera Bustamante, Yue Chang, Myriam Charfeddine, Tonny Chen, Xiaofu Chen, Manon Kirstine Chevalier, Shirisha Reddy Chintaguntla, Choon Hock Cho, Christian Christensen, Karl Ursø Christensen

Reading material:

Chapter 4, Chapter 5



Lecture Schedule

1 Introduction

31 January: C1

Data: Feature extraction, and visualization

2 Data, feature extraction and PCA

7 February: C2, C3

3 Measures of similarity, summary statistics and probabilities

14 February: C4, C5

4 Probability densities and data visualization

21 February: C6, C7

Supervised learning: Classification and regression

5 Decision trees and linear regression

28 February: C8, C9

6 Overfitting, cross-validation and Nearest Neighbor

7 March: C10, C12 (Project 1 due before 13:00)

7 Performance evaluation, Bayes, and Naive Bayes

14 March: C11, C13

8 Artificial Neural Networks and Bias/Variance

21 March: C14, C15

9 AUC and ensemble methods

28 March: C16, C17

Unsupervised learning: Clustering and density estimation

10 K-means and hierarchical clustering

11 April: C18

11 Mixture models and density estimation

18 April: C19, C20 (Project 2 due before 13:00)

12 Association mining

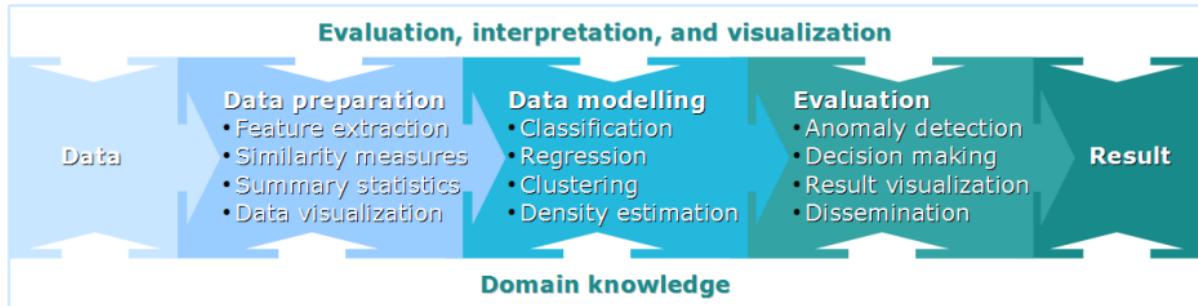
25 April: C21

Recap

13 Recap and discussion of the exam

2 May: C1-C21

Online 24/7 help: Discussion Forum/Piazza
Streaming & Videos: <https://panopto.dtu.dk/>
Online exercises: MS Teams



Learning Objectives

- Compute measures of similarity/dissimilarity (L_p distance, cosine distance, etc.)
- Understand basics of probability theory and stochastic variables in the discrete setting
- Understand probabilistic concepts such as expectations, independence and the Bernoulli distribution
- Understand the maximum likelihood principle for repeated binary events

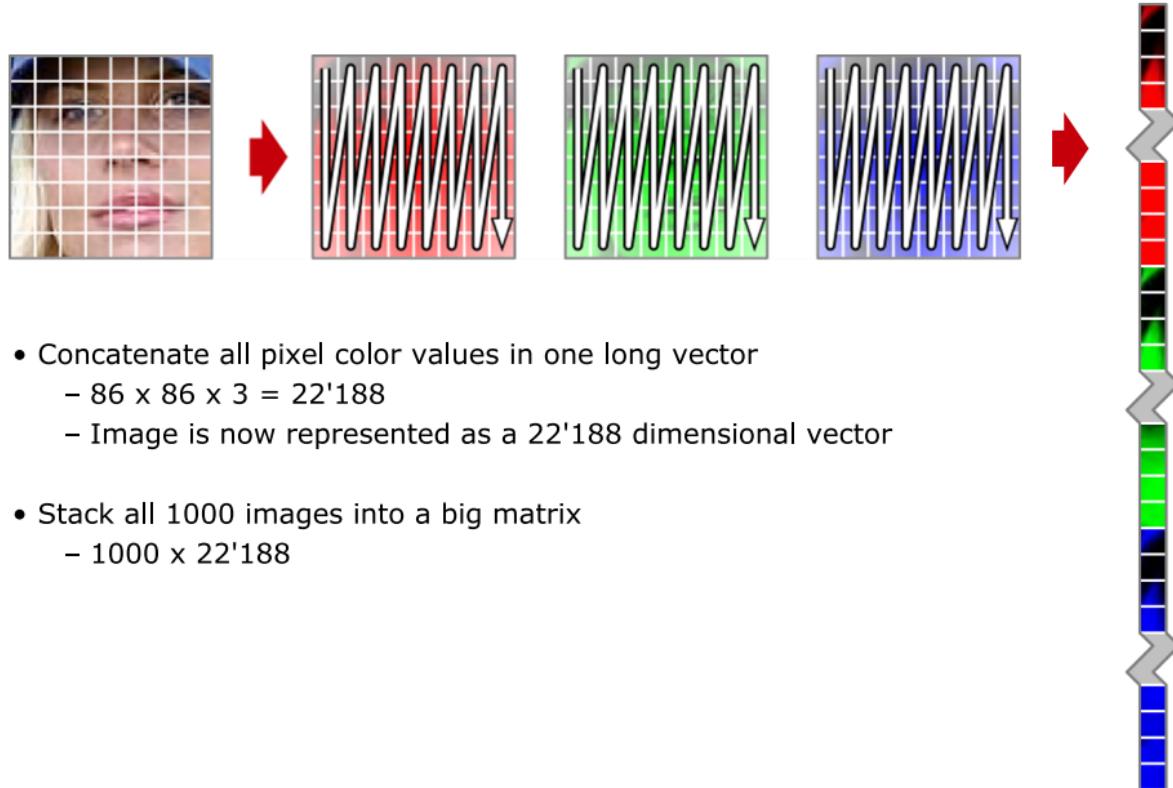
PCA recap: Principal component analysis on images



- 1000 images, 86 x 86 pixels, 3 RGB intensities

Tamara Berg "Faces in the wild"

Pre-processing



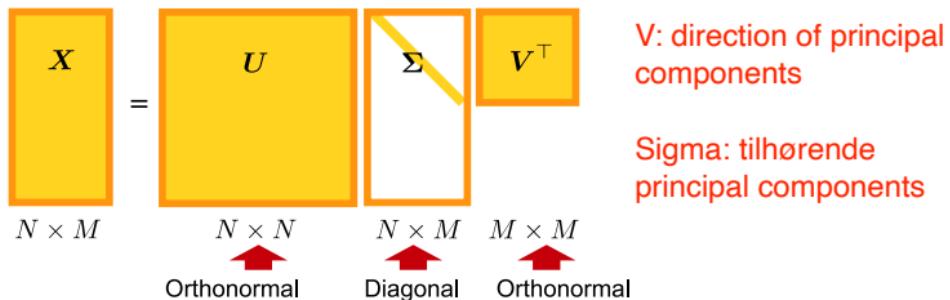
Principal component analysis (PCA)

1. Subtract the mean

- Consider dividing with variance; use 1-out-of-K coding for nominal attributes

2. Compute the singular value decomposition (SVD)

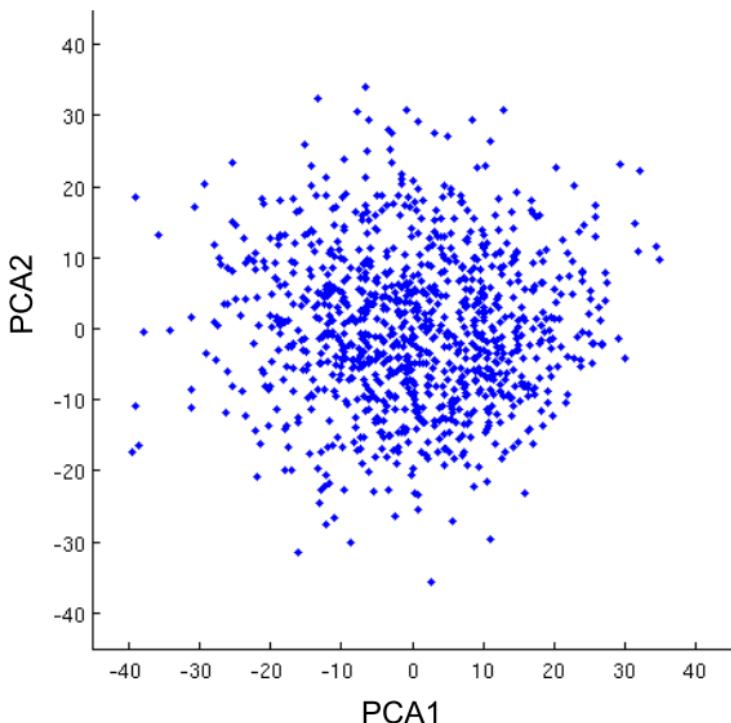
- Orthogonal linear transformation
- Transforms data to a new coordinate system
 - Greatest variance along the first axis (first column of V)
 - Second greatest variance along the second axis



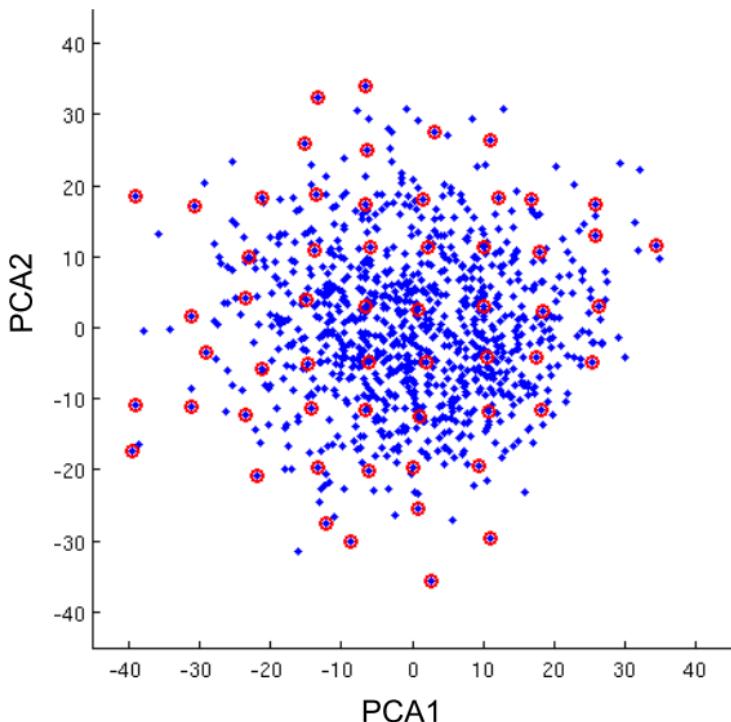
• Plot data in the transformed coordinate system

- Corresponds to looking at data from an angle where it is most spread out

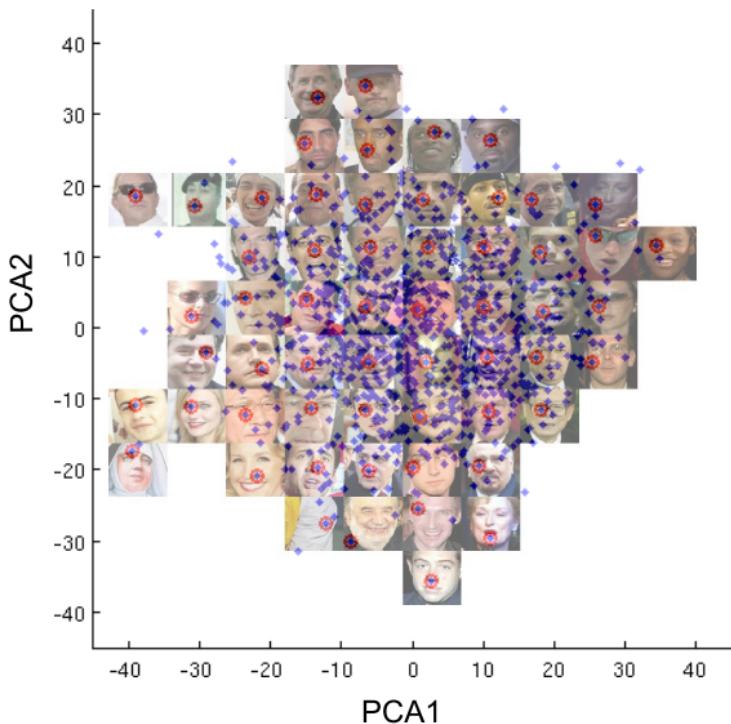
PCA on face images



PCA on face images

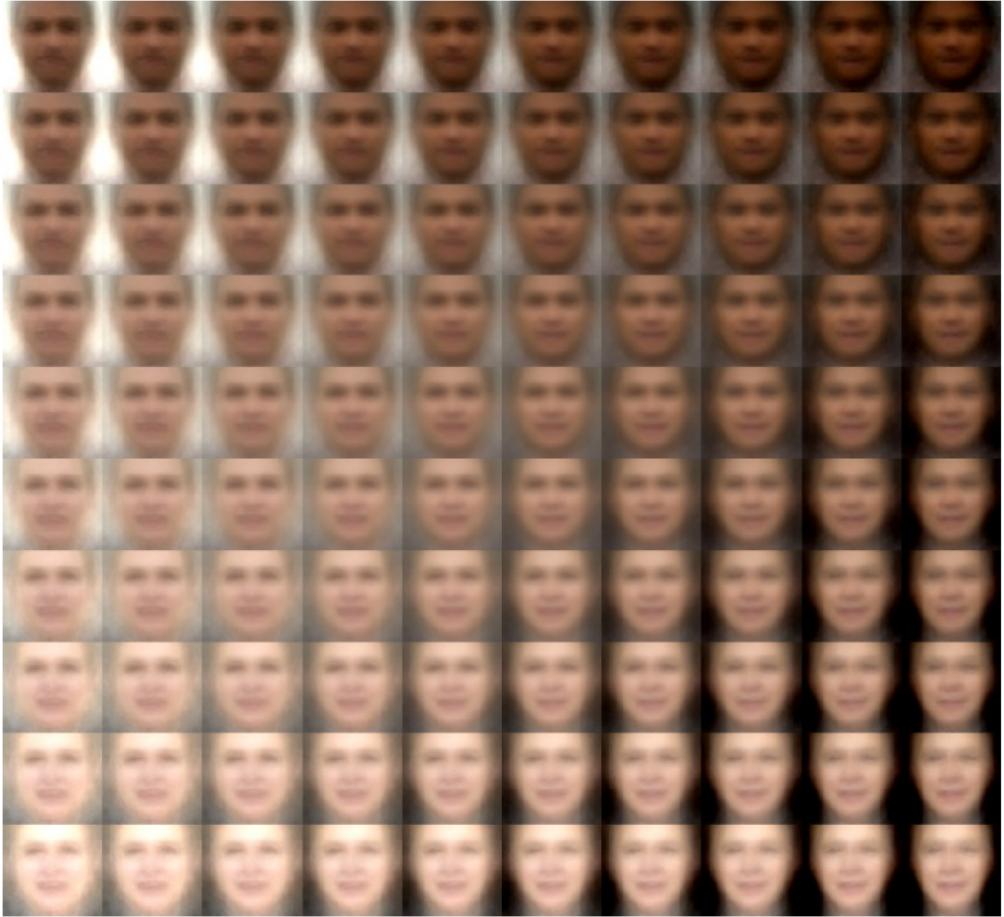


PCA on face images



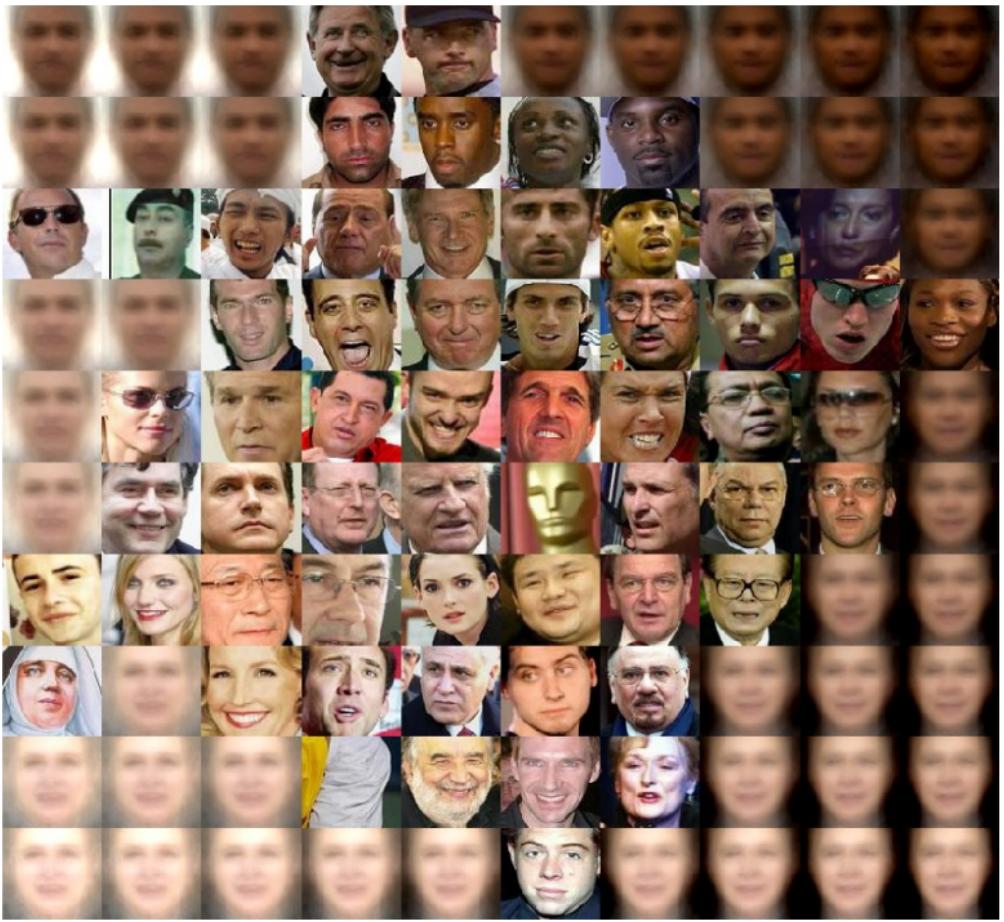


- What information do the two principal axes capture?





What information do the two principal axes capture?



Similarity / Dissimilarity measures

Similarity $s(x, y)$ Often between 0 and 1. Higher means more similar

Dissimilarity $d(x, y)$ Greater than 0. Lower means more similar.

Classification Classify a document as having the same topic y as the document it is **most similar/least dissimilar** to.

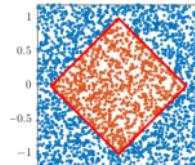
Outlier detection The observation most **dissimilar** to all other observations is an outlier



Dissimilarity measures

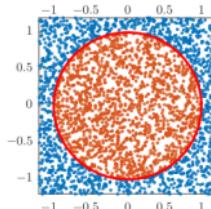
- General Minkowsky distance (p -distance) $d_p(\mathbf{x}, \mathbf{y}) = \left(\sum_{j=1}^M |x_j - y_j|^p \right)^{\frac{1}{p}}$
- One-norm ($p = 1$)

$$d_1(\mathbf{x}, \mathbf{y}) = \sum_{j=1}^M |x_j - y_j|$$



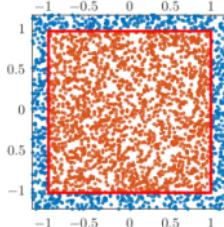
- Euclidean ($p = 2$)

$$d_2(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{j=1}^M (x_j - y_j)^2}$$



- Max-norm distance ($p = \infty$)

$$d_\infty(\mathbf{x}, \mathbf{y}) = \max \{|x_1 - y_1|, |x_2 - y_2|, \dots, |x_M - y_M|\}$$



Usage: Regularization and alternative optimization targets. For instance, $p = \infty$ is very affected by outliers, $p = 1$ much less so.

Similarity measures

$$x = \begin{bmatrix} 0 \\ 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} \quad y = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$$

K : Total number of attributes
 f_{00} : Number of attributes where $x_k = y_k = 0$
 f_{11} : Number of attributes where $x_k = y_k = 1$

Simple Matching Coefficient (SMC)

$$\text{SMC}(x, y) = \frac{f_{00} + f_{11}}{K}$$

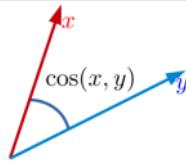
+ Symmetric

Jaccard Coefficient

$$J(x, y) = \frac{f_{11}}{K - f_{00}}$$

+ Positive matches

Cosine similarity



$$\cos(x, y) = \frac{x^\top y}{\|x\| \|y\|}$$

+ Positive matches
+ Document length

<-
Euklidisk
2-norm

Extended Jaccard coefficient

$$\text{EJ}(x, y) = \frac{x^\top y}{\|x\|^2 + \|y\|^2 - x^\top y}$$

Also defined for continuous data

Quiz 1, similarity measures

Calculate the simple matching coefficient, Jaccard, cosine and extended jaccard similarity between customer 1 and customer 2 in the market basket data below.

ID	Bread	Soda	Milk	Beer	Diaper
1	1	1	1	0	0
2	0	1	1	0	1

K : Total number of attributes

f_{00} : Number of attributes where $x_k = y_k = 0$

f_{11} : Number of attributes where $x_k = y_k = 1$

Which of the following statements are true?

- A. $\text{SMC}(o_1, o_2) = \frac{3}{5}$, $J(o_1, o_2) = \frac{1}{2}$, $\cos(o_1, o_2) = \frac{2}{3}$,
- B. $\text{SMC}(o_1, o_2) = \frac{3}{5}$, $J(o_1, o_2) = \frac{3}{4}$, $\cos(o_1, o_2) = \sqrt{\frac{2}{3}}$,
- C. $\text{SMC}(o_1, o_2) = \frac{2}{5}$, $J(o_1, o_2) = \frac{1}{3}$, $\cos(o_1, o_2) = \frac{2}{3}$,
- D. $\text{SMC}(o_1, o_2) = \frac{2}{5}$, $J(o_1, o_2) = \frac{1}{3}$, $\cos(o_1, o_2) = \sqrt{\frac{2}{3}}$,
- E. Don't know.

$$\text{SMC}(x, y) = \frac{f_{00} + f_{11}}{K}$$

$$J(x, y) = \frac{f_{11}}{K - f_{00}}$$

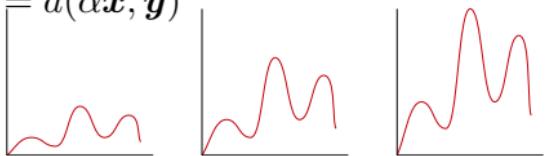
$$\cos(x, y) = \frac{x^\top y}{\|x\|_2 \|y\|_2}$$

$$\text{EJ}(x, y) = \frac{x^\top y}{\|x\|_2^2 + \|y\|_2^2 - x^\top y}$$

Invariance

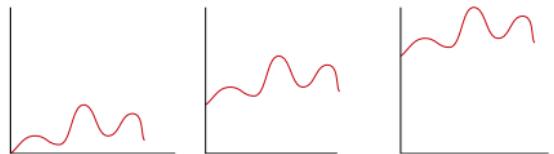
Scale invariance

$$d(\mathbf{x}, \mathbf{y}) = d(\alpha \mathbf{x}, \mathbf{y})$$



Translation invariance

$$d(\mathbf{x}, \mathbf{y}) = d(\alpha + \mathbf{x}, \mathbf{y})$$



General invariances

$$d(\boxed{6}, \boxed{2}) = d(\boxed{6}, \boxed{2})$$

Transformations

Standardization: Ensure a single attribute will not dominate:

$$\tilde{x}_{ik} = \frac{x_{ik} - \hat{\mu}_k}{\hat{\sigma}_k}$$

Example:

- **Number of children** ~ 0-5
- **Age** ~ 0-100 years
- **Annual income** ~ 0-50.000 €

Combining heterogeneous attributes Transform measures and combine

$$s_{\text{Edu.}} = \text{SMC}(x_{\text{Edu.}}, y_{\text{Edu.}})$$

$$s_{\text{Age.}} = a (a + d_1(x_{\text{Age.}}, y_{\text{Age.}}))^{-1}, \quad a = 1$$

$$s(x, y) = \frac{1}{2} (s_{\text{Edu.}} + s_{\text{Age.}})$$

Example:

- **Age:** Continuous
- **Education:** Binary
 - Primary (yes/no)
 - Secondary (yes/no)
 - Tertiary (yes/no)

Weighting Attributes have different importance

$$s(x, y) = 0.99s_{\text{Edu.}} + 0.01s_{\text{Age.}}$$

Empirical statistics

Given two samples $x_1, x_2, \dots, x_N \in \mathbb{R}$ and $y_1, y_2, \dots, y_N \in \mathbb{R}$:

- Empirical mean

$$\hat{\mu} = \bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$$

- Empirical variance

$$\hat{s} = \text{vár}[x] = \frac{1}{N-1} \sum_{i=1}^N (x_i - \hat{\mu})^2$$

- Empirical covariance

$$\text{côv}[x, y] = \frac{1}{N-1} \sum_{i=1}^N (x_i - \hat{\mu}_x)(y_i - \hat{\mu}_y)$$

- Empirical standard deviation

$$\hat{\sigma} = \text{std}[x] = \sqrt{\hat{s}}$$

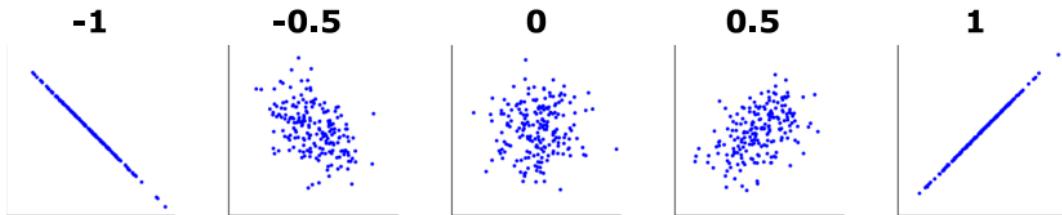
Correlation

- Measure of degree of linear relationship

$$\text{corr}[x, y] = \frac{\text{cov}[x, y]}{\hat{\sigma}_x \hat{\sigma}_y}$$

- A correlation of **1** or **-1** means there is a perfect linear relation

$$x_k = ay_k + b$$



Quantiles

Given N observations of an attribute $x_1, x_2, \dots, x_N \in \mathbb{R}$.

Quantiles describe the *points* that divide the underlying distribution into intervals that are equally probable:

- The one 2-quantile (**median**) divides the distribution in two intervals.
- The three 4-quantiles (**quartiles**) divides the distribution in four intervals.
- The 99 100-quantiles (**percentiles**) divides the distribution in 100 intervals.

The **median** is the same as the 2nd quartile or the 50th percentile.

E.g., we can (approximately) find the **median** by

- Sort the observations in ascending order $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(N)}$

$$\text{median}[x] = \begin{cases} x_{\left(\frac{N+1}{2}\right)} & \text{if } N \text{ is odd} \\ \frac{1}{2} \left(x_{\left(\frac{N}{2}\right)} + x_{\left(\frac{N}{2}+1\right)} \right) & \text{if } N \text{ is even.} \end{cases}$$

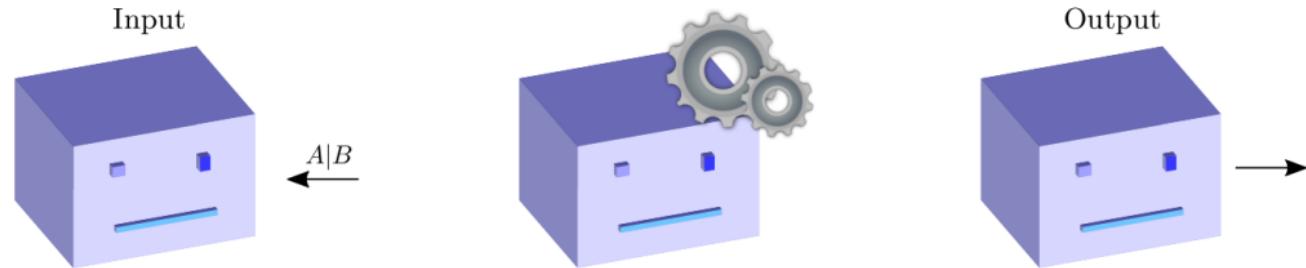
Probabilities

Pragmatically: A big part of AI is dealing with **uncertainty** and **incomplete information**. Probabilities is the formal framework for doing so.

Algorithmically: **If** an image belongs to a particular category is a discrete event. The **probability** it belongs to a category is continuous. Algorithmically, easier to optimize continuous quantities.

Convenience: There are boiler-plate ideas for transforming a **probabilistic** assumption into an algorithm (maximum likelihood).

Probabilities



Assuming B is true, how plausible is it A is true?

Assuming B , the plausibility of A is
(low / medium / high / certain)

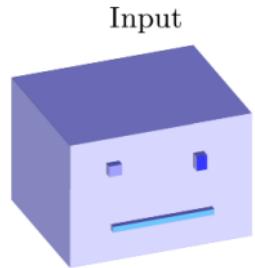
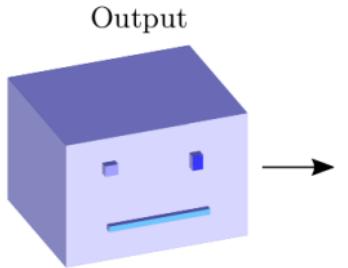
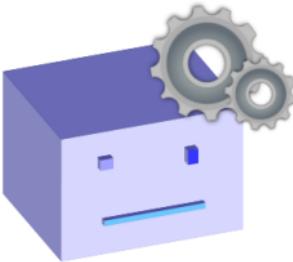
We reason about a proposition A in light of evidence B :

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

The degree-of-belief that A is true given B is accepted as true is at a level x

- A number between 0 and 1
- A and B are always binary (true/false) propositions
- Represents a *state of knowledge*

Probabilities: Trial example

 $A|B$ 

Assuming B is true, how plausible is it A is true?

Assuming B , the plausibility of A is
(low / medium / high / certain)

G : *The accused is guilty*

E_1 : *A car similar to his was seen at the crime scene.*

E_2 : *A large sum of money was found in his posession*

E_3 : *His fingerprints was found at the door of the bank.*

Probabilities express states-of-knowledge

$$E \equiv E_1 \text{ and } E_2 \text{ and } E_3$$

$$P(G|E) > P(G|E_2)$$

Binary propositions

A binary proposition is a statement which is either true or false (we might not know, but someone with complete knowledge would)

A : *In 49 BCE, Caesar crossed the Rubicon*

B : *Acceleration sensor 39 measures more than 0.85*

C : *Patient 901 has high cholesterol*

Propositions can be combined with **and**, **or** and **not**:

$AB \equiv$ True if A and B are both true

$A + B \equiv$ True if either A or B are true

$\bar{A} \equiv$ True if A is false

We define two special propositions which is always **true/false**:

1 : *A proposition which is always true*

0 : *A proposition which is always false*

...and the following identities: $A1 = A$, $A + \bar{A} = 1$, $\bar{\bar{A}} = A$ and

$$A(B_1 + B_2 + \cdots + B_n) = AB_1 + AB_2 + \cdots + AB_n$$

Quiz 2, Probabilities

Assume we define the following 4 boolean variables.

R_1 : Handed in report 1

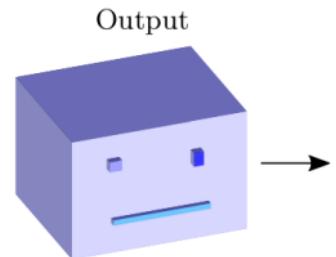
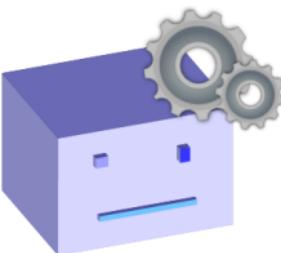
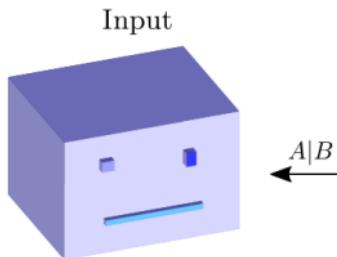
R_2 : Handed in report 2

R_3 : Handed in report 3

F : Student failed 02450

How would you express the probability of the statement:

If a student hand in report 1, 2 and 3, the chance of passing 02450 is greater than 90%?



Assuming B is true, how plausible is it A is true?

Assuming B , the plausibility of A is
(low / medium / high / certain)

Rules of probability

The sum rule: $P(A|C) + P(\bar{A}|C) = 1$

The product rule: $P(AB|C) = P(B|AC)P(A|C)$



Interpretation:

$P(A|B) = 0$ (*interpretation: given B is true, A is certainly false*)

$P(A|B) = 1$ (*interpretation: given B is true, A is certainly true*)

We also use the shorthand:

$$P(A|1) = P(A)$$

$$p(A) + P(\bar{A}) = 1$$

$$p(AB) = P(A|B)P(B)$$

Remarkably, this is the mathematical basis for this course

Marginalization and Bayes' theorem

Sum rule $P(A|C) + P(\bar{A}|C) = 1$

Product rule $P(AB|C) = P(B|AC)P(A|C)$

$$\begin{aligned} P(B|C) &= P(B|C) \left[P(A|BC) + P(\bar{A}|BC) \right] = P(AB|C) + P(\bar{A}B|C) \\ &= P(B|AC)P(A|C) + P(B|\bar{A}C)P(\bar{A}|C). \end{aligned}$$

$$\begin{aligned} P(A|BC) &= \frac{P(B|AC)P(A|C)}{P(B|C)} \\ &= \frac{P(B|AC)P(A|C)}{P(B|AC)P(A|C) + P(B|\bar{A}C)P(\bar{A}|C)}. \end{aligned}$$

DNA



Bayes theorem

$$P(A|BC) = \frac{P(B|AC)P(A|C)}{P(B|AC)P(A|C) + P(B|\bar{A}C)P(\bar{A}|C)}$$

Crimes may be solved by matching crime-scene DNA to DNA in a database

- If the two samples are from the same person, a DNA test will always give a positive match
- If the DNA are from different persons, DNA will incorrectly give a positive match one time out of a million

A crime is committed in Racoon City by an unidentified male. Assume all 8000 possible perpetrators undergo a DNA test, and suppose the DNA test gives a positive result for George. What is the chance George is guilty?



G : George is guilty, D : There was a positive DNA match

Exclusive and exhaustive events

A_1 : The side \square face up.

A_2 : The side \square face up.

A_3 : The side \diamond face up.

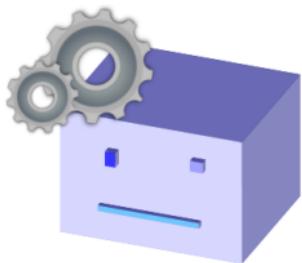
A_4 : The side \diamond face up.

A_5 : The side \heartsuit face up.

A_6 : The side \heartsuit face up.

- When no two propositions can be true at the same time, they are said to be **mutually exclusive**: $A_i A_j = 0$ for $i \neq j$
- Consider any two events A and B

$$P(A + B) = P(A) + P(B) - P(AB)$$



- In general, for n mutually exclusive events

$$P(A_1 + A_2 + \dots + A_n) = \sum_{i=1}^n P(A_i)$$

- A set of events is **exhaustive** if one has to be true: $A_1 + \dots + A_n = 1$. Then:

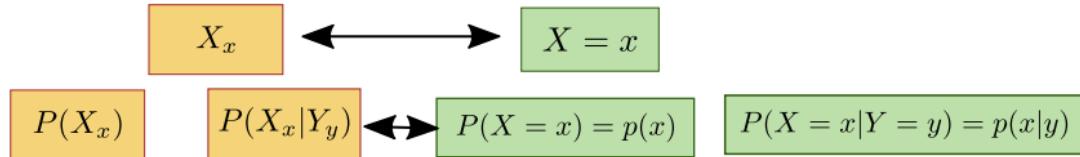
$$\sum_{i=1}^n P(A_i) = P(A_1 + A_2 + \dots + A_n) = 1$$

Stochastic variables

- Often, we will measure numerical quantities (number of children, age of a patient, etc.)
- Suppose a quantity X (number of children) takes a value $x = 3$. We can write this as the binary event X_3 and in general:

$X_x : \{\text{The binary event that } X \text{ is equal to the number } x\}$

- Stochastic variable simplify this notation by the definition:



Sum rule $P(A|C) + P(\bar{A}|C) = 1$

Product rule $P(AB|C) = P(B|AC)P(A|C)$

Marginalization

$$P(A|C) = P(A|BC)P(B|C) + P(A|\bar{B}C)P(\bar{B}|C)$$

Bayes theorem

$$P(A|BC) = \frac{P(B|AC)P(A|C)}{P(B|AC)P(A|C) + P(B|\bar{A}C)P(\bar{A}|C)}$$

Sum rule $\sum_i P(x_i|z_k) = 1$

Product rule $p(x_i, y_j|z_k) = p(x_i|y_j, z_k)p(y_j|z_k)$

Marginalization $p(x_i|z_k) = \sum_j p(x_i|y_j, z_k)p(y_j|z_k)$

Bayes theorem $p(y_j|x_i, z_k) = \frac{p(x_i|y_j, z_k)p(y_j|z_k)}{\sum_j p(x_i|y_j, z_k)p(y_j|z_k)}$

Quiz 3, Avila bible (Fall 2018)

$p(\tilde{x}_2, \tilde{x}_{10} y)$	$y = 1$	$y = 2$	$y = 3$
$\tilde{x}_2 = 0, \tilde{x}_{10} = 0$	0.19	0.3	0.19
$\tilde{x}_2 = 0, \tilde{x}_{10} = 1$	0.22	0.3	0.26
$\tilde{x}_2 = 1, \tilde{x}_{10} = 0$	0.25	0.2	0.35
$\tilde{x}_2 = 1, \tilde{x}_{10} = 1$	0.34	0.2	0.2

Table 1: Probability of observing particular values of \tilde{x}_2 and \tilde{x}_{10} conditional on y .

We will consider a dataset based on the Avila bible. We wish to predict the copyist ($y = 0, 1, 2$) of a bible based on the two typographic attributes *upperm* and *mr/is*. We suppose the attributes have been binarized such that *upperm* corresponds to $\tilde{x}_2 = 0, 1$ and *mr/is* to $\tilde{x}_{10} = 0, 1$. Suppose the probability for each of the configurations of \tilde{x}_2 and \tilde{x}_{10} conditional on the copyist y are as given in Table 1. and the prior probability of

the copyists is

$$p(y = 1) = 0.316, p(y = 2) = 0.356, p(y = 3) = 0.328.$$

Using this, what is then the probability an observation was authored by copyist 1 given that $\tilde{x}_2 = 1$ and $\tilde{x}_{10} = 0$?

- A. $p(y = 1|\tilde{x}_2 = 1, \tilde{x}_{10} = 0) = 0.25$
- B. $p(y = 1|\tilde{x}_2 = 1, \tilde{x}_{10} = 0) = 0.313$
- C. $p(y = 1|\tilde{x}_2 = 1, \tilde{x}_{10} = 0) = 0.262$
- D. $p(y = 1|\tilde{x}_2 = 1, \tilde{x}_{10} = 0) = 0.298$
- E. Don't know.

Sum rule $\sum_i p(x_i|z_k) = 1$

Product rule $p(x_i, y_j|z_k) = p(x_i|y_j, z_k)p(y_j|z_k)$

Marginalization $p(x_i|z_k) = \sum_j p(x_i|y_j, z_k)p(y_j|z_k)$

Bayes theorem $p(y_j|x_i, z_k) = \frac{p(x_i|y_j, z_k)p(y_j|z_k)}{\sum_j p(x_i|y_j; z_k)p(y_j|z_k)}$

Independence

Independent: $p(x_i, y_j) = p(x_i)p(y_j)$

Conditionally independent given z_k : $p(x_i, y_j | z_k) = p(x_i | z_k)p(y_j | z_k)$

Expectations

Expectation: $\mathbb{E}[f] = \sum_{i=1}^N f(x_i)p(x_i).$ (2)

mean: $\mathbb{E}[x] = \sum_{i=1}^N x_i p(x_i),$ Variance: $\text{Var}[x] = \sum_{i=1}^N (x_i - \mathbb{E}[x])^2 p(x_i).$ (3)

Example: Uniform probability

$$p(x_i) = \frac{1}{N}$$

$$\mathbb{E}[f] = \frac{1}{N} \sum_{i=1}^N f(x_i)$$

$$\mathbb{E}[x] = \frac{\sum_{i=1}^N x_i}{N}$$

$$\text{Var}[x] = \frac{1}{N} \sum_{i=1}^N (x_i - \mathbb{E}[x])^2$$

Densities and models

- In machine learning, we want to learn a parameter from data
- Models of the data which use parameters are how we do that
- We build models out of simpler building blocks (distributions (discrete variables see chapter 5) and densities (continuous variables see chapter 6)). In this course we will use four:

Bernoulli distribution

The Categorical distribution

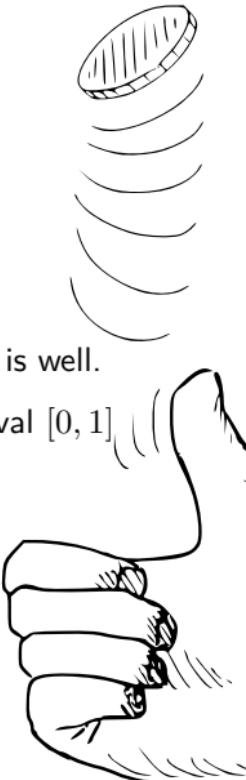
The Beta density

The Multivariate normal density

The Bernoulli distribution

- Let $b = 0, 1$ denote a binary event.
- For instance,
 - $b = 0$ corresponds to heads, and $b = 1$ to tails, or
 - $b = 0$ corresponds to a person being ill, and $b = 1$ that a person is well.
- The probability of b is expressed using a parameter θ in the unit interval $[0, 1]$

Bernoulli distribution: $p(b|\theta) = \theta^b(1 - \theta)^{1-b}$.





The Bernoulli distribution, repeated events

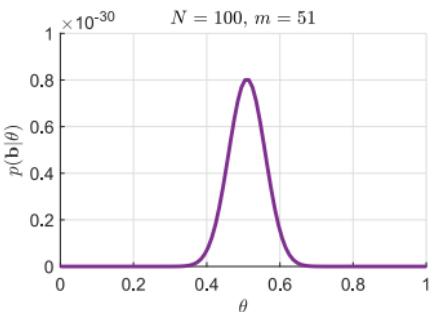
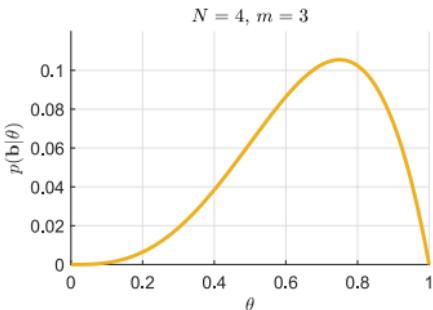
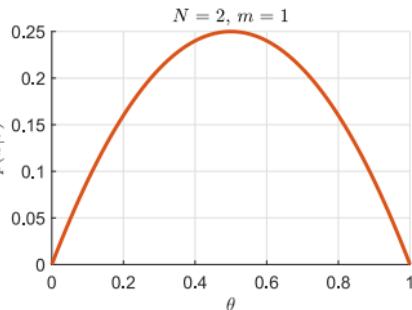
Conditional independence $p(x_i, x_j | z_k) = p(x_i | z_k)p(x_j | z_k)$

- Suppose we observe a sequence b_1, \dots, b_N of Bernoulli (binary) events.
- For instance, for N patients we record whether person 1 is ill or well ($b_1 = 0$ or $b_1 = 1$) and up to whether patient N is ill or well ($b_N = 0$ or $b_N = 1$)
- When we **know** θ (the chance a person is well or ill), the events are **independent**

Bernoulli distribution: $p(b|\theta) = \theta^b(1-\theta)^{1-b}.$

$$\begin{aligned} p(b_1, \dots, b_N | \theta) &= \prod_{i=1}^N p(b_i | \theta) = \prod_{i=1}^N \theta^{b_i}(1-\theta)^{1-b_i} = \theta^{\sum_{i=1}^N b_i}(1-\theta)^{N-\sum_{i=1}^N b_i} \\ &= \theta^m(1-\theta)^{N-m}, \quad m = b_1 + b_2 + \dots + b_N \end{aligned}$$

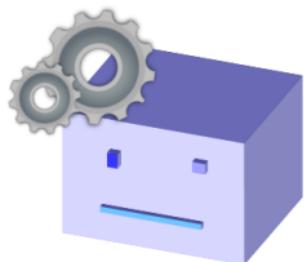
The Bernoulli distribution, maximum likelihood



$$p(b_1, \dots, b_N | \theta) = \theta^m (1 - \theta)^{N-m}$$

An idea for selecting θ^* is **Maximum likelihood**

$$\theta^* = \arg \max_{\theta} p(b_1, \dots, b_N | \theta)$$



The value of θ according to which the data is most plausible

Resources

<https://bayes.wustl.edu> Classical textbook which treats probabilities as states-of-knowledge and discuss many practical and philosophical issues (this book converted me to ML!)

(<https://bayes.wustl.edu/etj/prob/book.pdf>)

<https://02402.compute.dtu.d> A more in-depth description of summary statistics (see chapter 1) (<https://02402.compute.dtu.dk>)

<https://www.khanacademy.org> An excellent introduction to probability theory which we recommend as a go-to resource

(<https://www.khanacademy.org/math/statistics-probability/probability-library>)

<http://citeseerx.ist.psu.edu> A more in-depth discussion of Bayes in the court room (<http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=EF0328140036A4C668BB5B9FC76C9BE?doi=10.1.1.599.8675&rep=rep1&type=pdf>)