

Introduction to course 02457



Lars Kai Hansen, DTU Compute,
Course 02457 Sep 4, 2017



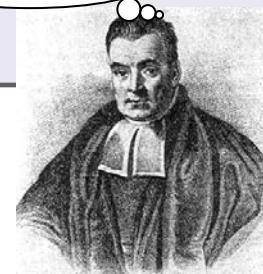
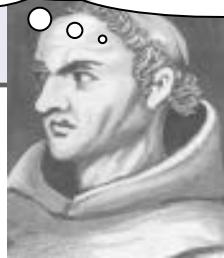
Technical University of Denmark

Outline

Do not multiply causes!

Multiply likelihood and prior..

$$P(a | b) = P(b | a)P(a) / P(b)$$



Introduction

What is machine learning?

Signal processing in the brain

Search engines

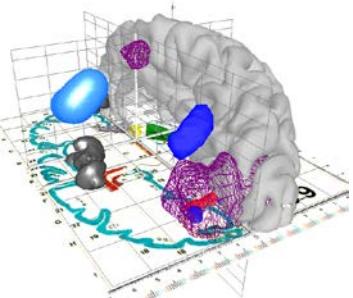
Course overview

What kinds of math are necessary?

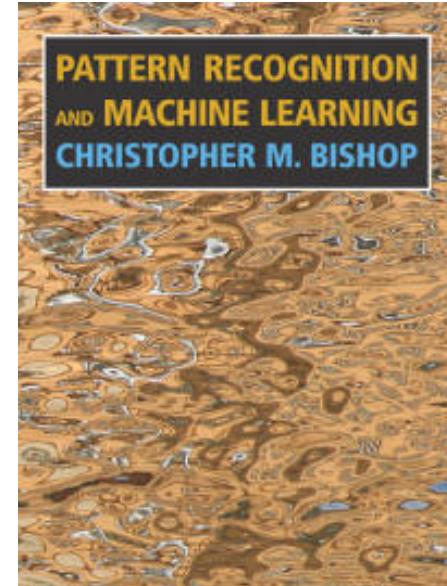
Probabilities, densities

Bayes theorem

Bayes decision theory

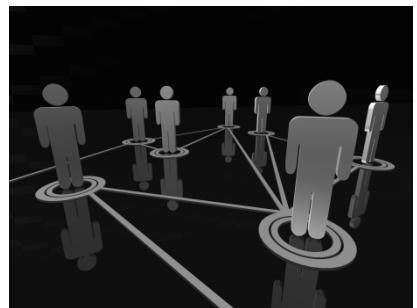


Textbook



Cognitive Systems

A Section in DTU Compute
Approx. 50 scientists, students,
staff



Social
computing



Ceci n'est pas un smartphone

Digital
media



Cognitive
Neuroscience



Statistical
machine
learning

Mobile
Systems

Neuro-
informatics



What is machine learning?

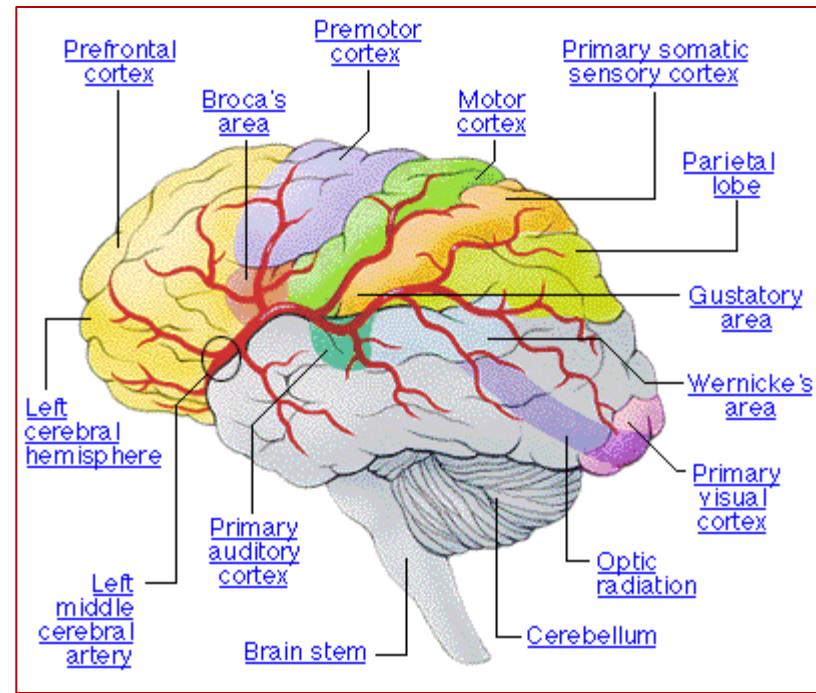
Abstraction of generalizable relations from empirical data

Statistical modeling in domains where we have no or little insight into the physics of the problem

Inspired by / interfacing with the brain – people involved!

Emphasis on

- Prediction
- Scaling
- Robustness
- Interpretability



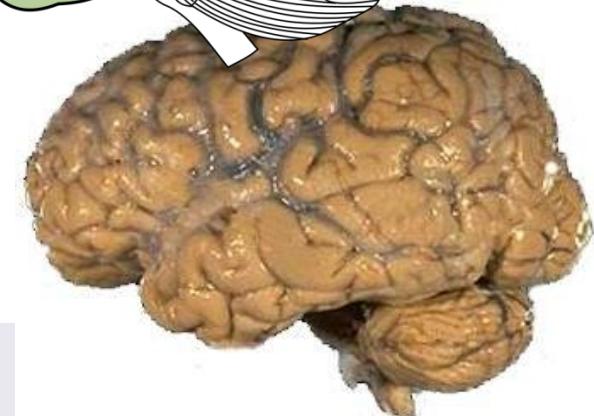
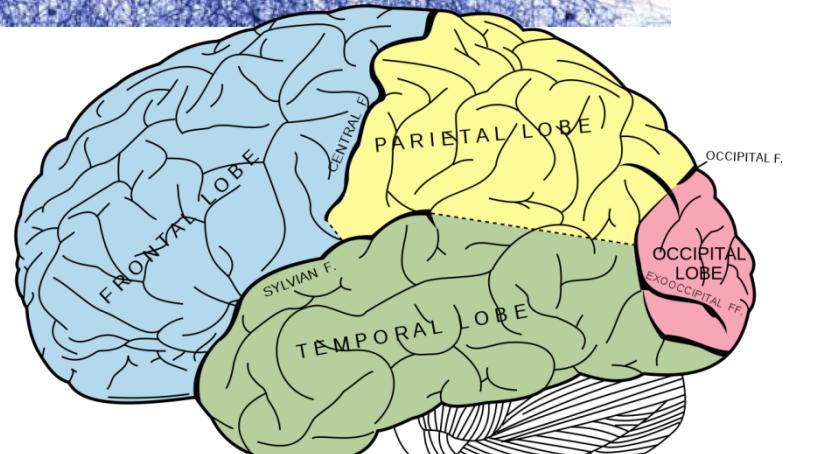
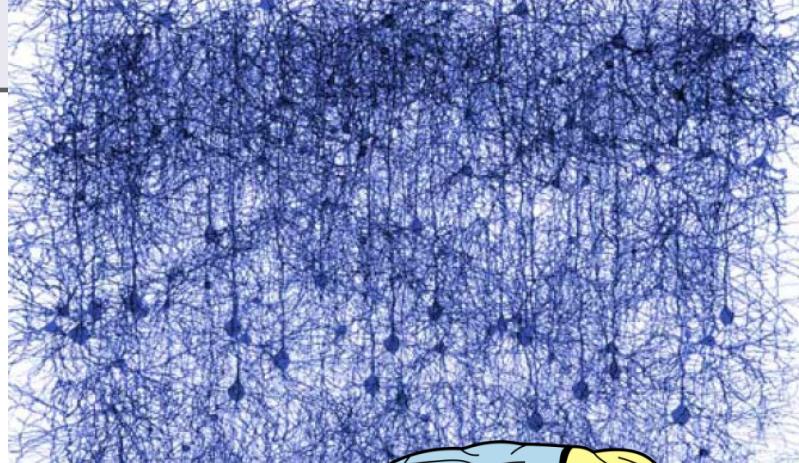
The brain design

Three design principles

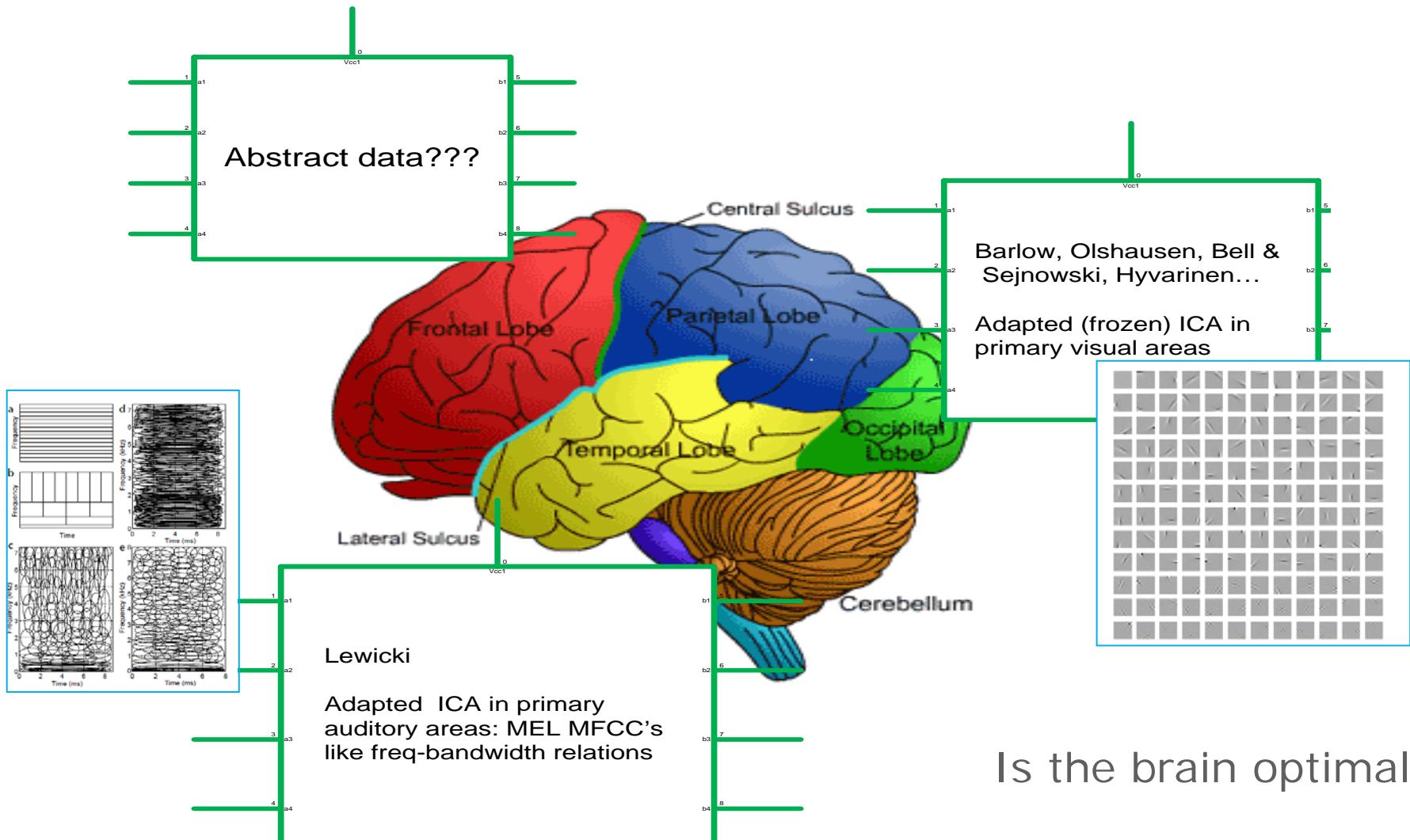
- i) Division of labor
- ii) Neural networks of simple computers
- iii) Learning – adaptivity -plasticity

Centers for vision, hearing, smell etc

The brain's "output": motor signals



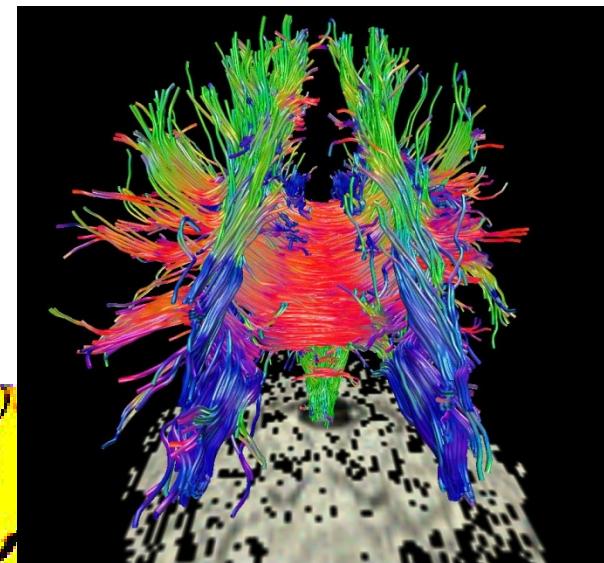
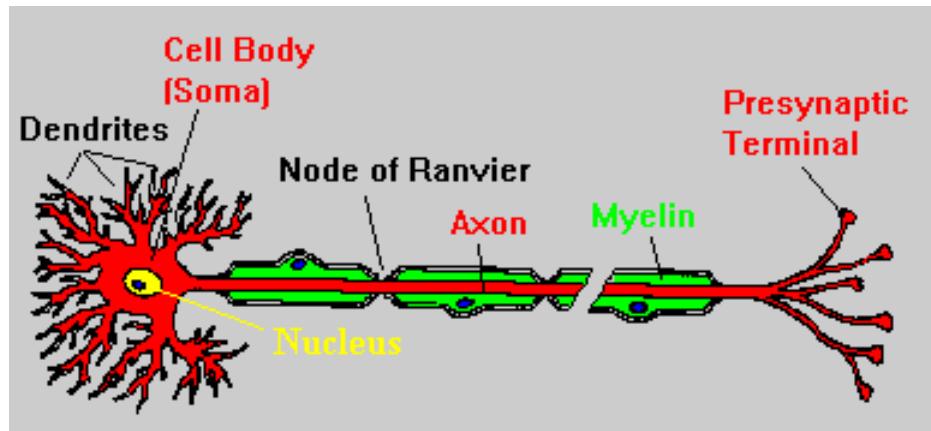
Division of labor: Understanding dependency structures in environment



Neural circuitry

Cortex consists of gray matter brain cells / neurons, modeled as simple computers ...1 bit?

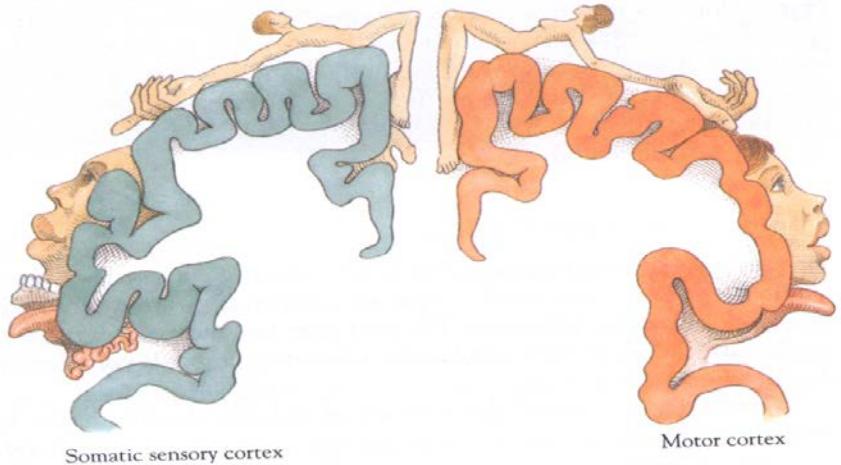
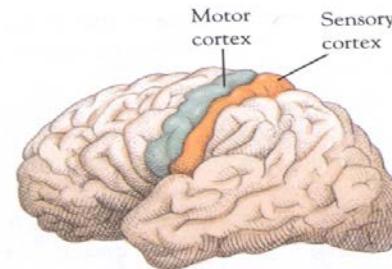
and white matter: wires connecting neurons at synapses (+/-)



Sensory input - Motor output

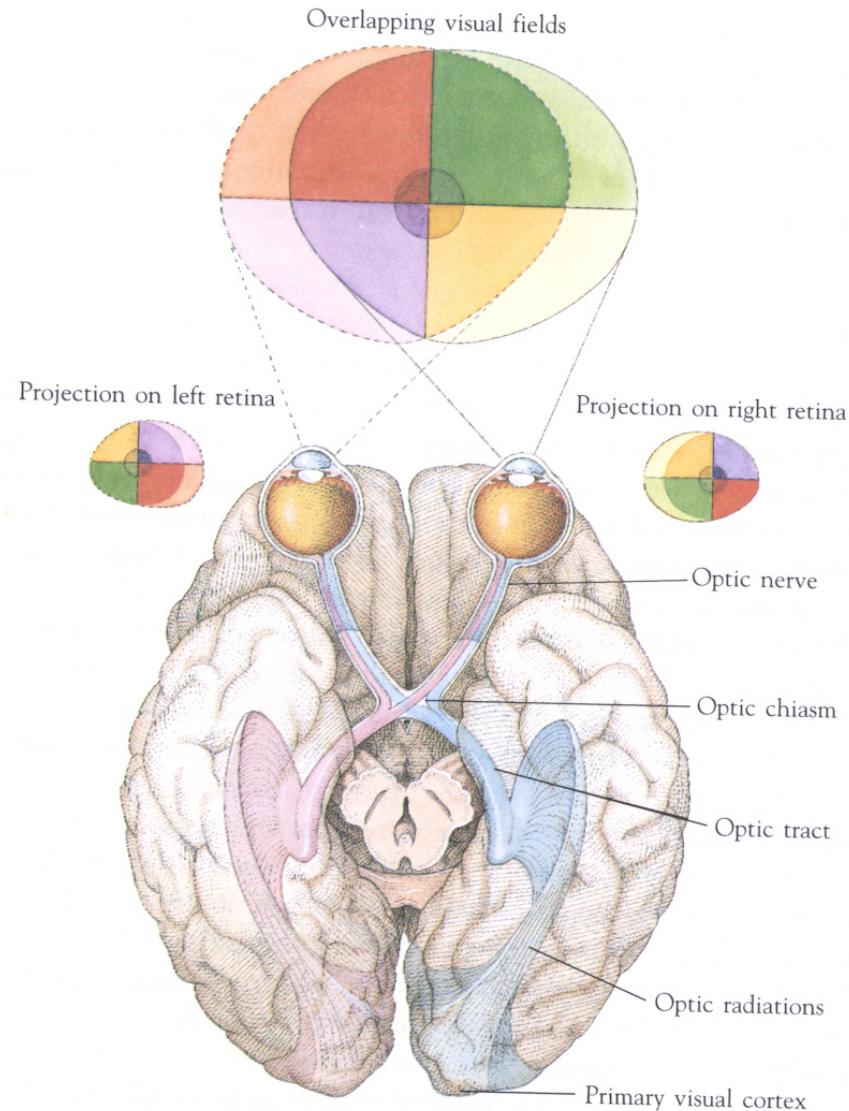
- Multiple sensory inputs from vision, ears, skin,...
- Muscles are activated by specific areas in the brain

CHAPTER ONE



Vision resides in the back of the human head

The "GPU" is a massively parallel computer

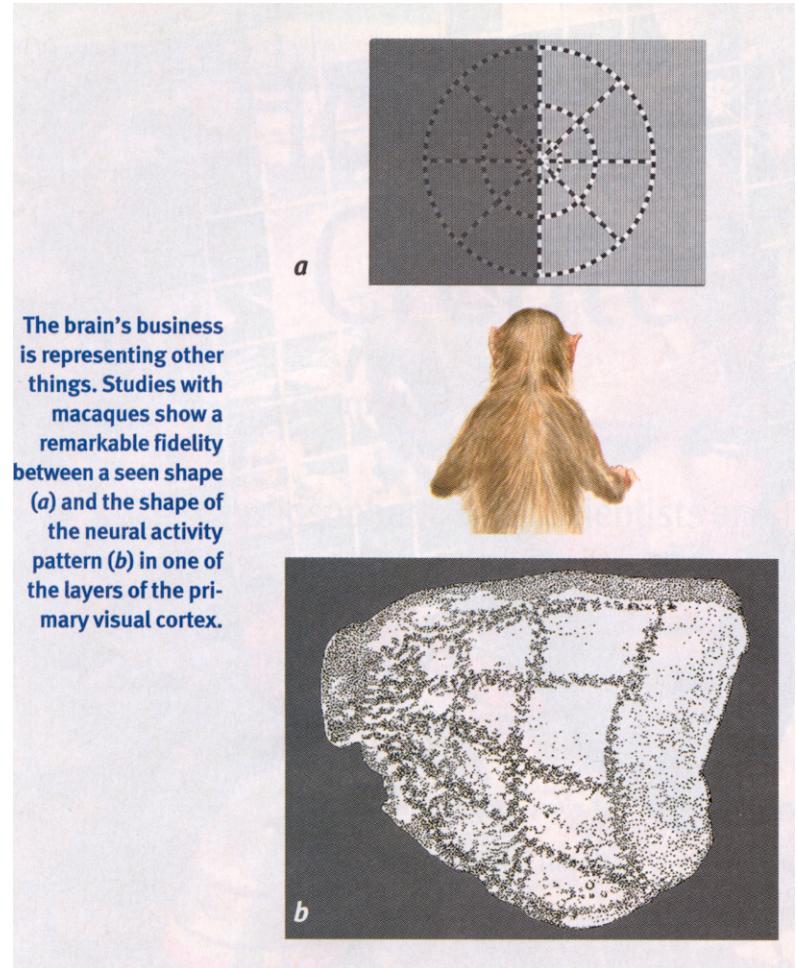


Visual cortex is adapted to natural stimuli

Visual cortex creates multiple maps: contrast, depth, color, shape,... “division of labor”

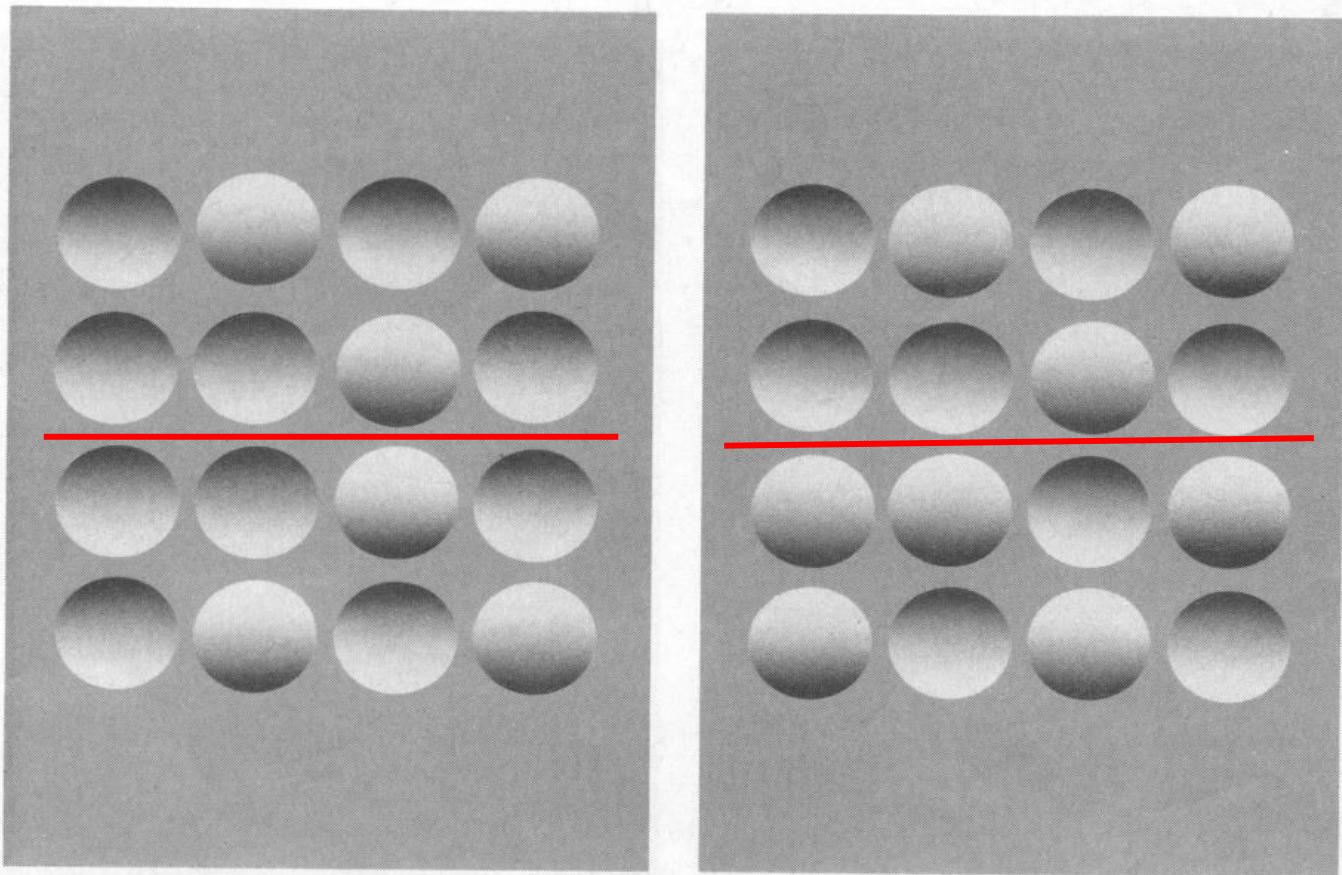
Visual cortex is plastic

Visual cortex is massively parallel



Vision software

The human brain's context setting determines visual interpretation



SYMMETRY PERCEPTION occurs after the brain extracts shape from shading. In the left-hand array spheres and cavities seem arranged symmetrically about a horizontal axis. But in two dimensions it is the right-hand array that is truly symmetrical.

Visual system hierarchy (Van Essen)

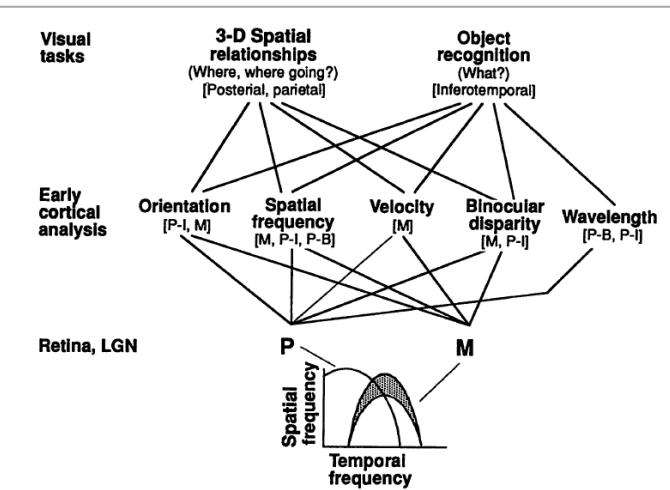
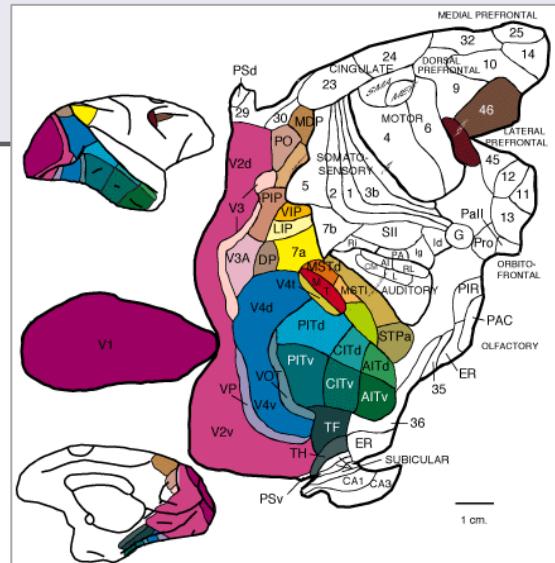
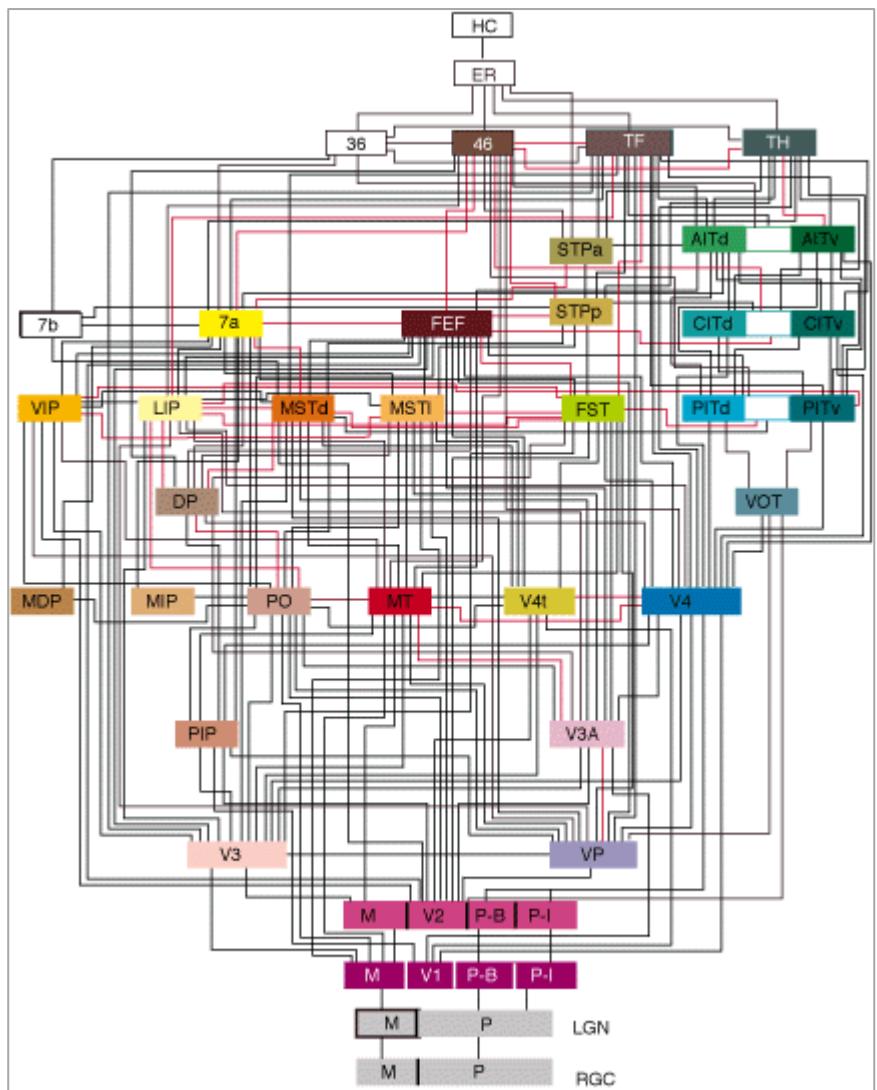
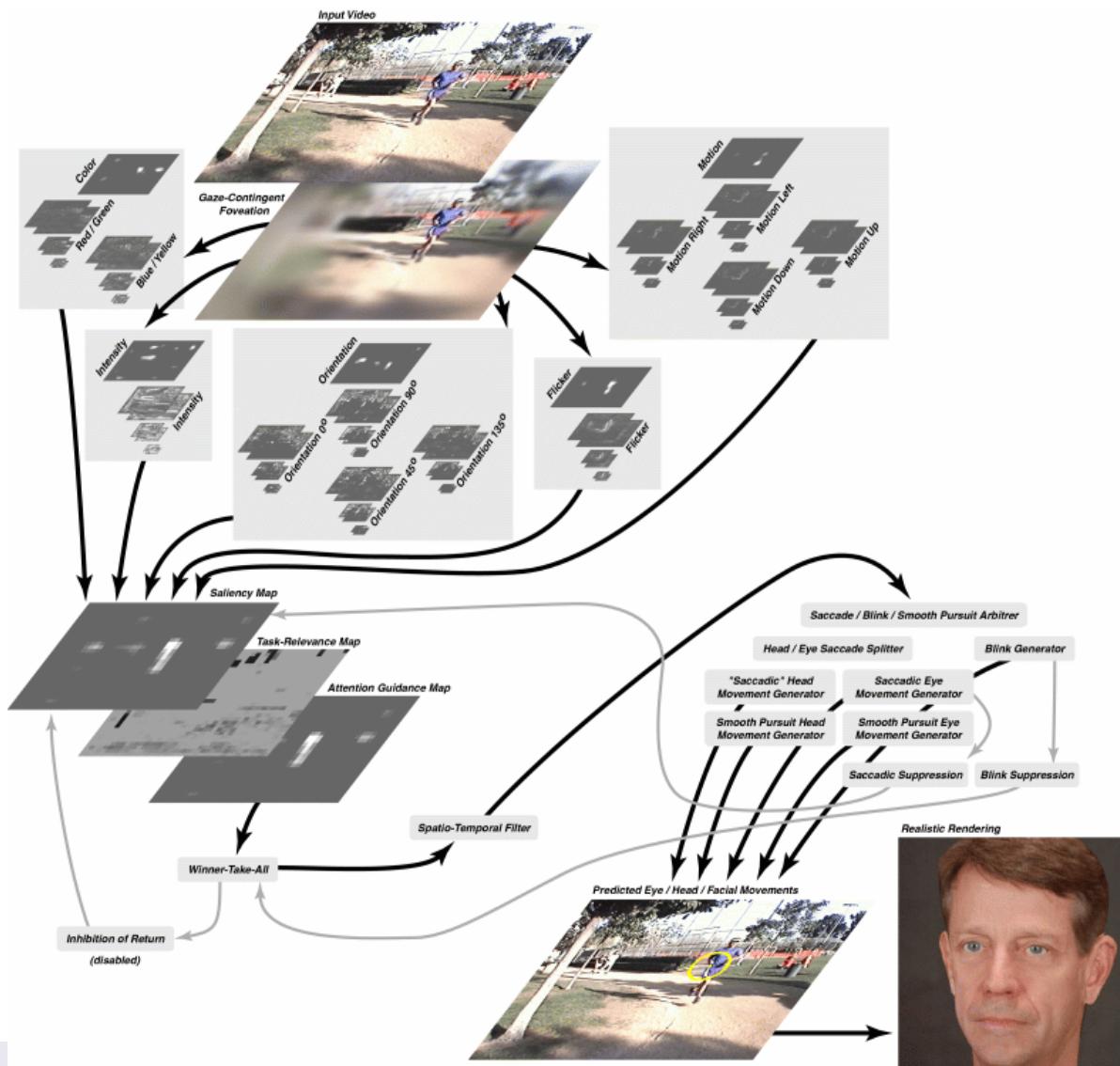


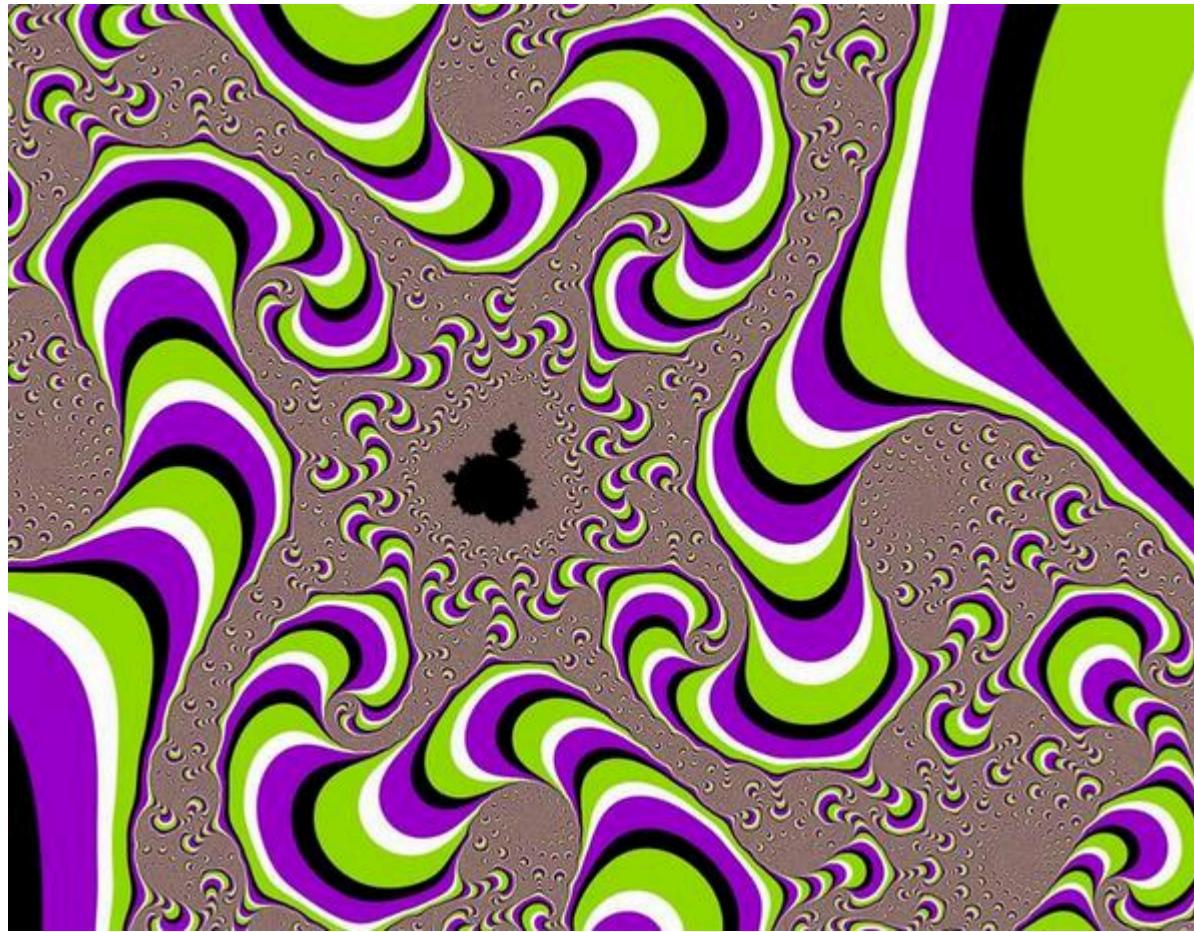
Fig. 3. Convergence and divergence in visual processing. Arrows represent major lines of information flow from subcortical P and M streams (bottom) to the selectivities represented among neurons at early stages of cortical analysis (middle) and from there to two general tasks of vision (top level). The hatched portion of the M cell curve represents their nonlinear component of processing. The processing streams associated with each property in the middle row are assigned on the basis of a high incidence of selectivity recorded physiologically (6, 7).

Felleman, D.J., and D.C. Van Essen. "Distributed hierarchical processing in the primate cerebral cortex." *Cerebral cortex* 1, no. 1 (1991): 1-47.

Itti, Dhavale & Pighin, Proc. SPIE, 2003



Damn you visual cortex! Other domains.. Cognitive biases



the vision

Computeren bliver så småt en del af os

Fremtidens computere og mobiltelefoner behøver ikke at ligge i tasker eller bukselommer. Om få år kan de sidde på armen, næsen og andre steder på kroppen eller ligefrem i øjet, vurderer eksperter.

TEKNOLOGI

CHRISTIAN GRØNBØT PETERSEN

Forsigtig, at du ikke længere behøver tage telefonen op af lommen. Det er denne ambition, der har drævet firmaet bag sagenværktøjen Google til at udvikle en intelligent computerbrille. I denne uge begyndte firmaet at sende de første 2.000 eksemplarer afbriller kaldet Google Glass på gaden til særligt udvalgte brugere, der har berett for at teste den nye teknologi.

Det er et sunt skridt i retning mod fremtidens kropstørre computere, som kan blive altidere for øvel den bærbare computer som de mobiltelefoner, der kendes i dag. Men brillerne er bare et blandt flere projekter, der er påværet fra lab-horrorerne ud til brugeren.

Google Glass er en brille uden ørygge og glas, men med en lille computer og en kamera monteret i siden af næller, som sender billeder og indber synseleter med informationer fra en smartphone.

«Vi forsøger ikke, siden at forsuduge, hvad der bliver der næste, iringenen gennem vil have», forklarer Googles danske taleslunde, Christine Sørensen, og tilføjer at vi som brugere af smartphones allerede er vant til at have teknologien nær på os.

«Med Glass bliver det bare et blik, der udskiller os».

Informationerne i Glass kan være alt fra kørersejledninger, smitter og Google-søgninger, og de kan variere, efter hvort man befinner sig. Billedet i brugerens smartphone tilde Google til en 25 mm med lidt skærm set fra 2,5 meters afstand. De hele opgave ved hjælp af stemmekom-

ce på DTU, venter utilmodigt på de brillen, hans forskningsafdeling har bestilt.

«Af holdet til maven, vi bruger mest på i dag, er næsteåndes computer, som Glass, naturligvis angivelse for smartphones, siger han.

Intelligent teknologi

Brillene er dog ikke nødvendigt den mest oplyste computer for almindelige mennesker at tage på, mener Hanne-Louise Johansen, som underviser på IT-universitetet i København og er ekspert i såkaldte wearable computers. Hun er også partner i et firma, som arbejder med at udvikle næsteåndes teknologi.

«En af fordele ved noget er, at den kommer mere diskret end eksempletvis Google Glass, som jeg synes, er meget overbausende, jeg vil ikke i hvert fald ikke selv bryde sig om at gå rundt med dem», siger hun.

Forsker fra University of Roeter i Southhampton, der også er en del af universitetet i Cambridge, har nu nemlig udviklet et materiale, som kan være fremtidens for teknologi.

Materiale er lavet på det ny type opfundne kulmfoldede græn, som ikke bare er meget stærkt, men også god til at lede strøm. Derfor kan den bruges til at have et, der indeholder computer og andre elektroniske dele.

«Det kan komme et enormt potentiale, siger Hanne-Louise Johansen og præger på fremtiden forskning og udvikling inden for intelligente materialer. Måndag kan døse med din.

Hier hjemme har firmaet HypoSafe sammentoed DTU-forskere været med til at udvikle diabetes-alarmet.

Ved hjælp af et lille implantat i huden kan patienter få besked om lav blodsockerniveau, før det bliver farlige. Og på universitetet i Arizona har man udviklet intelligente strømpe, der kan overvåge temperatur og blodtryk, så lagen ikke



The information you need when you need it

Support productivity and well-being Respect democratic living and privacy

virtuelle og virkelige», forklarer Kirsten Poulsen, som har et firma, der arbejder med at identificere disse mennesker.

Og det er nærp, hvad Google Glass forsøger at gøre, mener Lars Kai Hansen fra DTU.

Men Google er ikke de eneste på fejret. En del andre firmaer arbejder også på teknologi, der kan give brugeren informationer i et smartwatch, når man skal til arbejde, eller ved at installere Samsung har allerede bekraet, at de arbejder på et lignende projekt.

Siden er det den syriske firmalet Thalmic Labs, der udvikler en teknologi, som ved hjælp af enkle sensorer i hånden, elektroniskravat kan udlevere informationer i maskinerne og omstændigheder til digitale kommandoer. Armedamler kan med andre ord omstænde armen til en fjernbetjening, så et knippe eller vilt med fingrene kan standse musikken på en computer eller sætte for fjernsynet.

Forbud mod computerbriller

Den catalytiske professor Steve Mann havde i 1990'erne udviklet Google Glass, før den blev kendt, fordi han siden 2000 er en aktiv protestant på hjemmelavede computerbriller, han også kalder Glass. Og om i den syvste version minder om Google-brillerne.

Men i modsætning til Google er Mann mål med brillerne primært at forbedre læsernes syn. Hans briller kan se varme, der ved at udvikle øjeblikkelige øjeblikke, der gør folk i stand til at gæde næste gang, når de gæder næste gang.

Hansen, professor, DTU, kritiserer Google Glass, som han mener fylder folk med for meget information. Hanne-Louise Johansen er enig i den kritik.

«Det virker lidt, som om Google Glass forsøger at gøre for mange ting på én gang», siger han.

http://www.google.dk/#hl=da&rlz=1R2GFRE_daDK376&q=~jaguar&aq=f&aqi=g10&aql=&oq=&gs_rfai=&fp=511881e210c1f1b1

Søg Del Sidewiki Bogmærker Kontroller Oversæt Lars.K... Convert Select

Favorites Web Slice Gallery

~jaguar - Google-søgning

Nettet Billeder Kort Oversæt Blogs Indeks Gmail mere ▾

Lars.Kai.Hansen@gmail.com | Weboversigt | Indstillinger | Log ud

Google

~jaguar

Ca. 725.000.000 resultater (0,41 sekunder)

Søg Avanceret søgning

Jaguar.dk ☆
Salg af nye og brugte biler. Værksted, reservedele og tilbehør. Sitet indeholder oplysning om modeller, billeder og nyheder. Virtuelt showroom med butik og ...
Tilkørte biler - Nye biler på lager - Priser & udstyr
www.jaguar.dk/ - Cached - Lignende

TIGER - Det er for vildt! ☆
Tiger har altid noget du kan bruge - hele året rundt. Masser af sjove, nyttige ting inden for bolig, pleje, legetøj, dvd'er, kontor, krydderier og meget ...
Tigers katalog - Adresser og åbningstider - Job i Tiger - Hobby & fritid
www.tiger.dk/ - Cached - Lignende

jaguars.com - 2010 Season Tickets On Sale Now! ☆ - [Oversæt denne side]
The official team site with scores, news items, game schedule, and roster.
2010 Draft - Tickets - Message Board - TEAM
www.jaguars.com/ - Cached - Lignende

Billeder af ~jaguar - Rapporter billeder

Jaguar International - Jaguar International ☆ - [Oversæt denne side]
Our mission at Jaguar has been to create and build beautiful fast cars. The XK, XF, X-TYPE and now All New XJ bring the exhilaration of driving to life.
www.jaguar.com/ - Cached - Lignende

Sponsorerede links

Jaguar
Læs designerbriller kun kr. 1495
Upload dit foto og prøv dem online
www.louisnielsen.dk

Jaguar
Find ny brugt bil - ALD CarMarket!
kvalitetsbiler m. fejl fri FDM-tests
www.aldcarmarket.dk

Jaguar på bil.guide
Test og omtaler af Jaguar.
Alt i guides, sport, luksus & brugt
bil.guide.dk/Jaguar

Frisørsaks.dk
Danmarks største udvalg
af kvalitets frisørsakse.
www.frisorsaks.dk

Jaguar på Biltorvet
35.000 brugte biler fra hele landet
find din Jaguar på Biltorvet
www.biltorvet.dk

Nissan service
Autoservice til Nissan,

Done Internet | Protected Mode: Off 100% DA 16:08 18-05-2010

Google getting smarter - semantics

jaguar.dk - Jaguar i Danmark

Annonce www.jaguar.dk/ ▾

officiel importør af Jaguar - se alle modeller her!

Om Jaguar

Blav klogere på Jaguar & vores spændende historie.

Jaguar XF Sportbrake

Se den flotte Jaguar XF Sportbrake.
- Spændende og funktionelt design.

[Forside | Jaguar - Jaguar](#)

www.jaguar.dk/ ▾

Velkommen til Jaguars officielle hjemmeside i Danmark.

Jaguar

www.jaguar.com/ ▾ [Oversæt denne side](#)

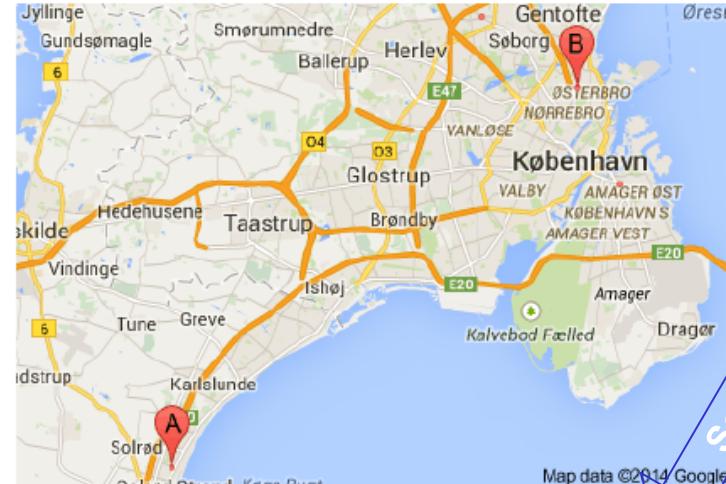
Official worldwide web site of Jaguar Cars. Directs users to pages tailored to country-specific markets and model-specific websites.

Billeder af jaguar



Rapportér billeder

[Flere billeder af jaguar](#)



[Kort over jaguar](#)

Se resultater om

Jaguar Cars

Jaguar Cars Limited er et britisk luksusbilmærke ejet af indiske Tata Motors med hovedkvarter i Browns ...



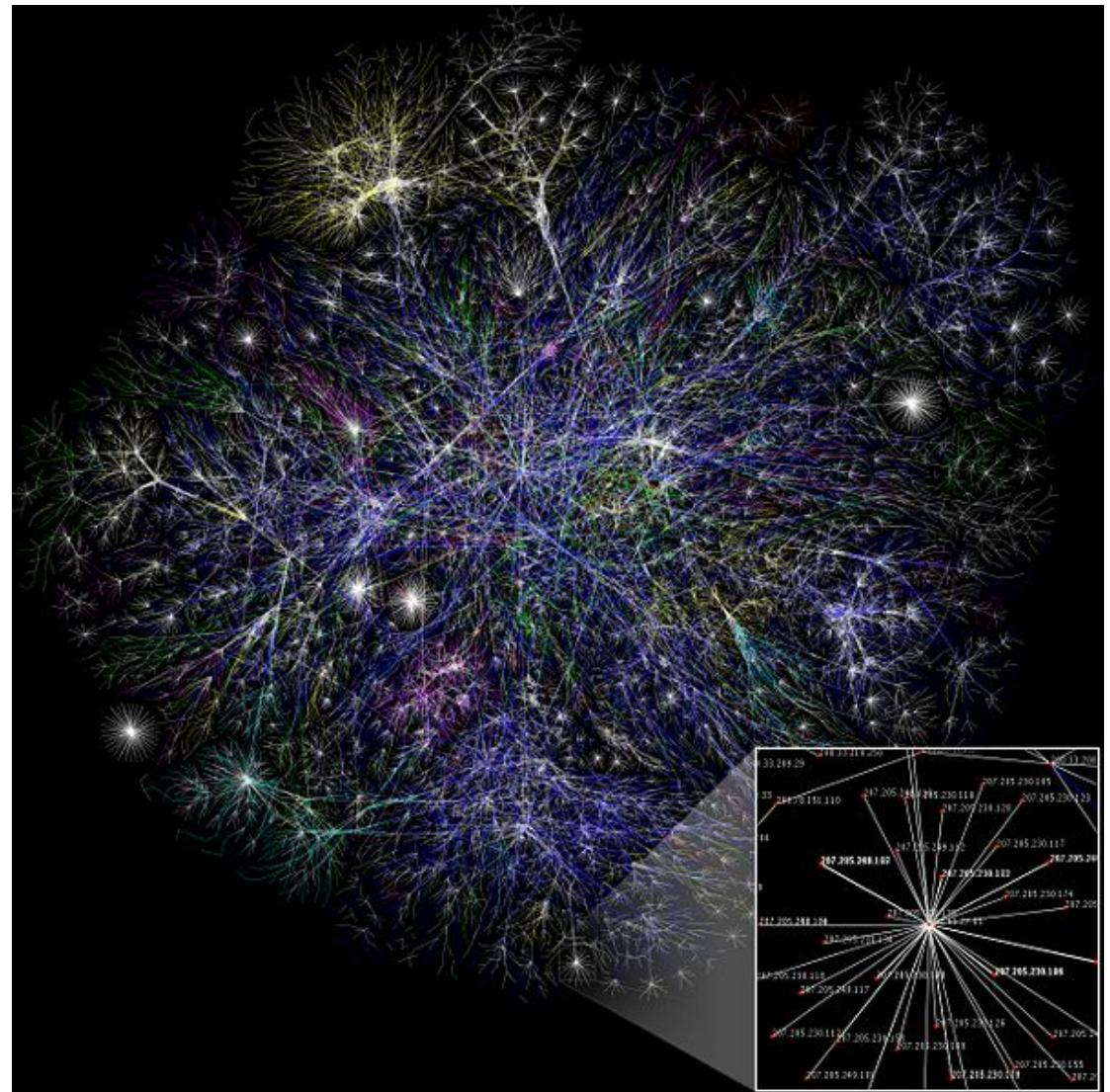
Jaguar

Jaguaren er et af de store kattedyr i kattefamilien. Den lever i Syd og Centralamerika og er tæt ...



... numerous specialized big data indexes for specialized search

Wikipedia (facts)
Facebook (friends)
Youtube (video)
Spotify (music)
Dating.dk (dates)
Twitter (news)



A search engine is a database of documents, an Index, and a retrieval function *Query -> Ranked retrieval*

Purpose

Return relevance ranked information in response to a *query*

Search related machine learning research issues

What should be indexed?what are the "words"?

Query expansion –query is a document (including context?)

Retrieval of documents and context information

Ranking of retrieval set

Relevance feedback, user modeling

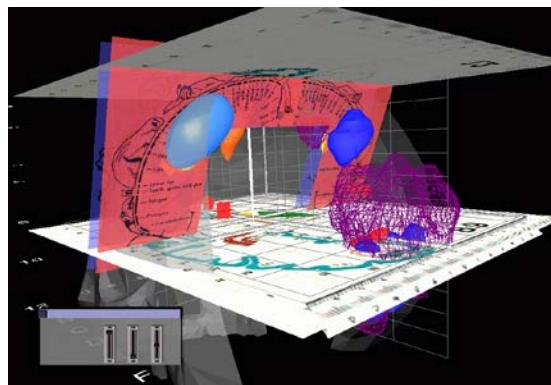
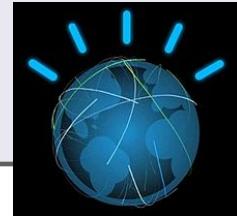
Organization of multi-dimensional responses

list, tree, groups etc

Maintenance issues:

Cleaning, updating, forecasting etc.

Search Apps



audio engines
MI Rocket (2006)
CastSearch (2007)
Muzeeker (2009)
CoSound (2014)

medical search engines
Brede Search neuroinformatics (2005)
FindZebra, diagnostic queries (2013)

A screenshot of the FindZebra In Press website, which compiles news articles from various sources about the search engine. The articles include coverage from the MIT Technology Review, The New York Times, The Telegraph, the guardian, NewScientist, and Smithsonian.com. Each article snippet provides a brief summary of how the search engine aids medical professionals in diagnosing rare diseases.

CNN Castsearch - Windows Internet Explorer

File Edit View Favorites Tools Help

Google "intelligent sound" matlab to Start Bogmærker 44 blokeret Kontroller Send til intelligent sound matlab Indstiller

CNN Castsearch

Trends : About

Search: schwarzenegger

Traditional Text Search

30/06/2006 23:00 Play segment Play file Transcription
 30/06/2006 14:00 Play segment Play file Transcription
 26/12/2006 05:00 Play segment Play file Transcription
 23/05/2006 10:00 Play segment Play file Transcription
 18/11/2006 13:00 Play segment Play file Transcription
 15/01/2007 13:00 Play segment Play file Transcription
 07/06/2006 11:00 Play segment Play file Transcription
 07/06/2006 10:00 Play segment Play file Transcription
 31/12/2006 03:00 Play segment Play file Transcription
 30/10/2006 01:00 Play segment Play file Transcription

Search by Expanded Query

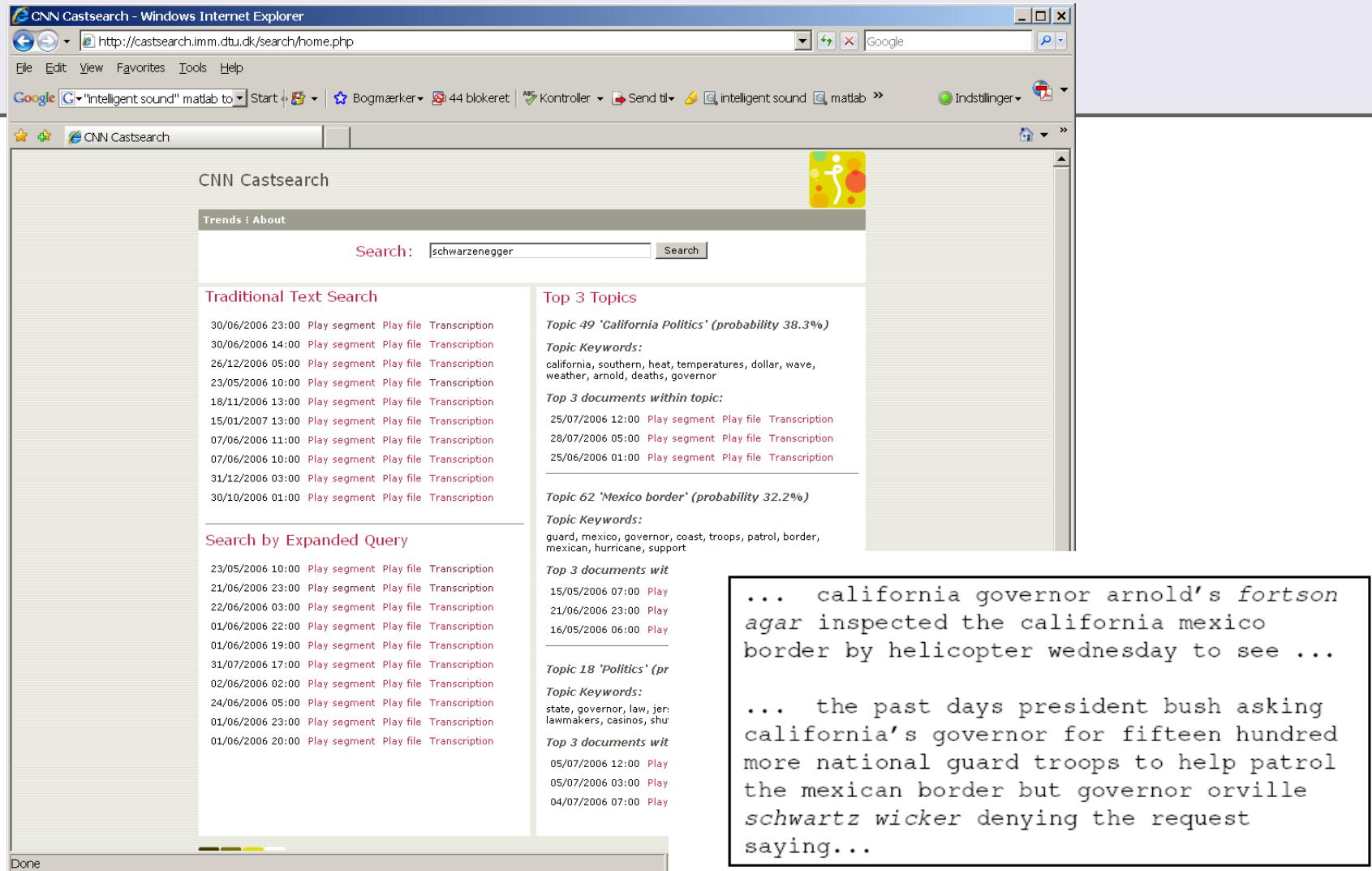
23/05/2006 10:00 Play segment Play file Transcription
 21/06/2006 23:00 Play segment Play file Transcription
 22/06/2006 03:00 Play segment Play file Transcription
 01/06/2006 22:00 Play segment Play file Transcription
 01/06/2006 19:00 Play segment Play file Transcription
 31/07/2006 17:00 Play segment Play file Transcription
 02/06/2006 02:00 Play segment Play file Transcription
 24/06/2006 05:00 Play segment Play file Transcription
 01/06/2006 23:00 Play segment Play file Transcription
 01/06/2006 20:00 Play segment Play file Transcription

Top 3 Topics

Topic 49 'California Politics' (probability 38.3%)
Topic Keywords:
 california, southern, heat, temperatures, dollar, wave, weather, arnold, deaths, governor
Top 3 documents within topic:
 25/07/2006 12:00 Play segment Play file Transcription
 28/07/2006 05:00 Play segment Play file Transcription
 25/06/2006 01:00 Play segment Play file Transcription

Topic 62 'Mexico border' (probability 32.2%)
Topic Keywords:
 guard, mexico, governor, coast, troops, patrol, border, mexican, hurricane, support
Top 3 documents wit
 15/05/2006 07:00 Play
 21/06/2006 23:00 Play
 16/05/2006 06:00 Play

Topic 18 'Politics' (pr
Topic Keywords:
 state, governor, law, jer, lawmakers, casinos, shu
Top 3 documents wit
 05/07/2006 12:00 Play
 05/07/2006 03:00 Play
 04/07/2006 07:00 Play



... california governor arnold's fortson agar inspected the california mexico border by helicopter wednesday to see ...

... the past days president bush asking california's governor for fifteen hundred more national guard troops to help patrol the mexican border but governor orville schwartz wicker denying the request saying...

Fig. 2. Two examples of the retrieved text for a query on 'schwarzenegger'.

castsearchimm.dtu.dk

CASTSEARCH - CONTEXT BASED SPEECH DOCUMENT RETRIEVAL

Lasse Lohilahti Mølgaard, Kasper Winther Jørgensen, and Lars Kai Hansen

Informatics and Mathematical Modelling
 Technical University of Denmark Richard Petersens Plads
 Building 321, DK-2800 Kongens Lyngby, Denmark

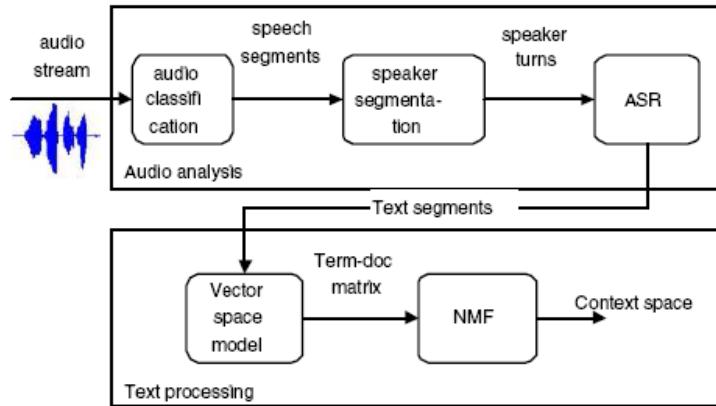
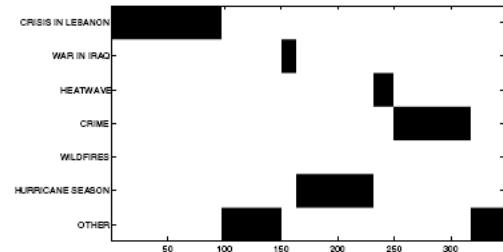
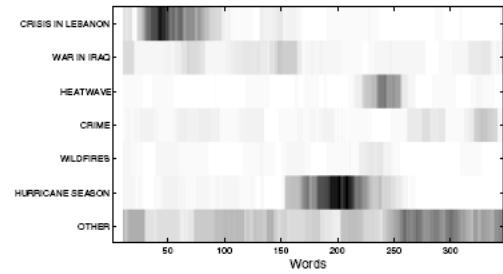


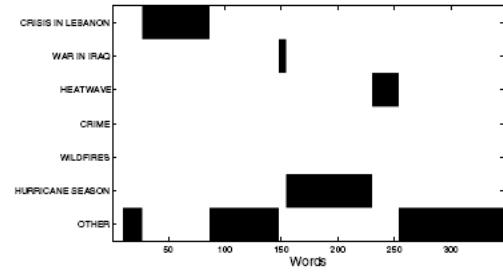
Fig. 1. The system setup. The audio stream is first processed using audio segmentation. Segments are then using an automatic speech recognition (ASR) system to produce text segments. The text is then processed using a vector representation of text and apply non-negative matrix factorization (NMF) to find a topic space.



(a) Manual segmentation.



(b) $p(k|d^*)$ for each context. Black means high probability.



(c) The segmentation based on $p(k|d^*)$.

Fig. 3. Figure 3(a) shows the manual segmentation of the news show into 7 classes. Figure 3(b) shows the distribution $p(k|d^*)$ used to do the actual segmentation shown in figure 3(c). The NMF-segmentation is in general consistent with the manual segmentation. Though, the segment that is manually segmented as 'crime' is labeled 'other' by the NMF-segmentation

From pipelines to deep networks

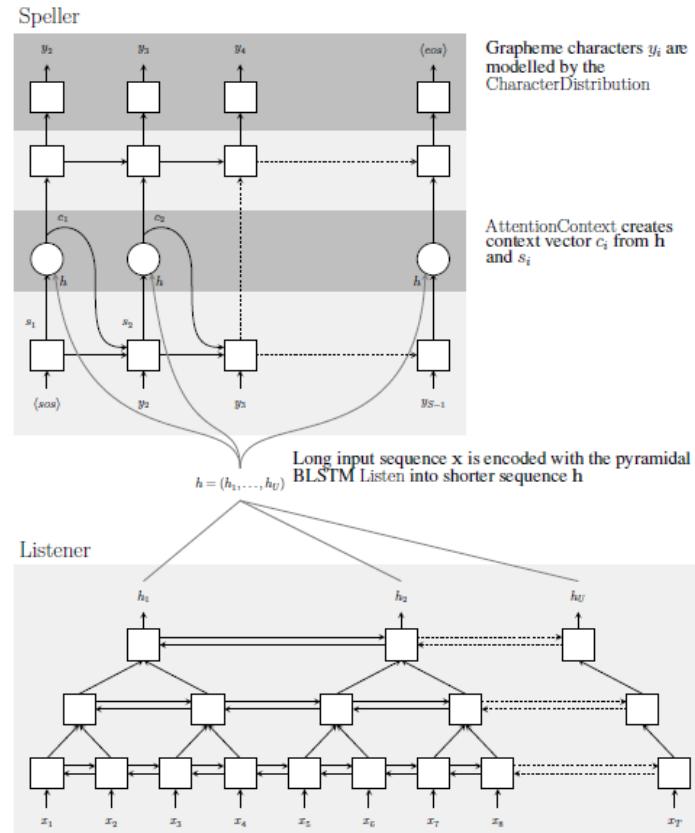
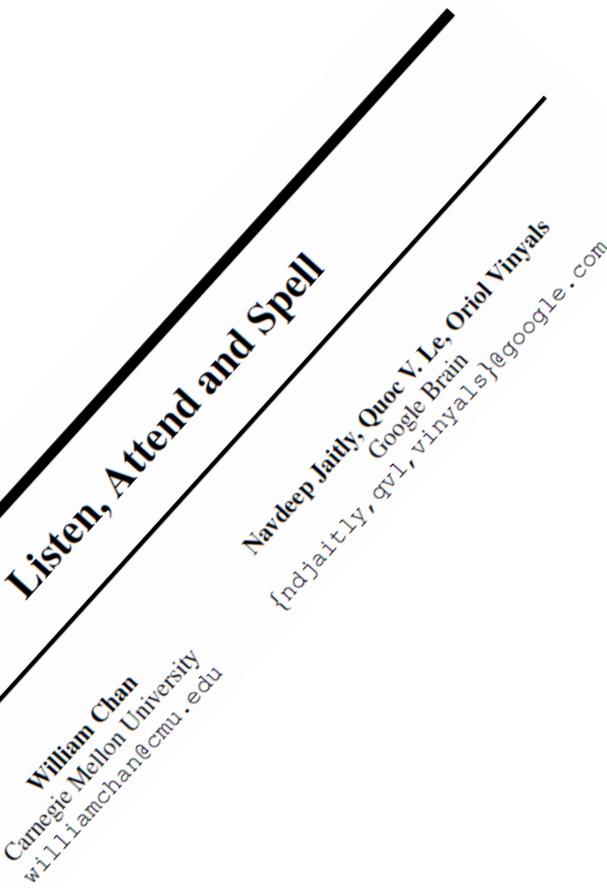


Figure 1: Listen, Attend and Spell (LAS) model: the listener is a pyramidal BLSTM encoding our input sequence x into high level features h , the speller is an attention-based decoder generating the y characters from h .

Machine learning: strategy for modeling objects in complex vector space representations (x)

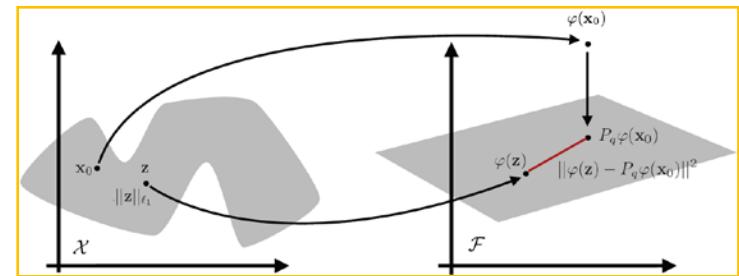
Measurements represented in vector spaces –
a point in a high-dimensional "feature space"

Simplest learning idea: object similarity \sim spatial proximity
–similarity measures / distance metrics become important,

More realistic: proximity in complex manifolds representing the physical constraints
- "long range" dependency

May need complex statistical models and lots of data (deep networks / big data)
to discover such representations from data

...e.g. "objects" in image space occupy highly complex manifolds determined by
(approximate) invariances



Unsupervised learning in vector spaces

Identification of structure

- Preprocessing/ dim redux
- Missing data
- De-noising
- Novelty / outlier detection

Methods

- Clustering
- Signal separation linear/non-linear

$$p(\mathbf{x} | \theta)$$

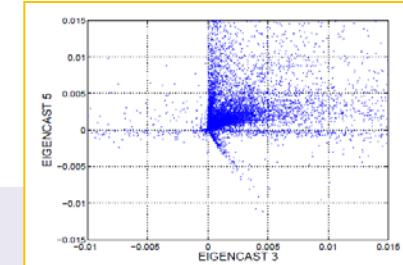
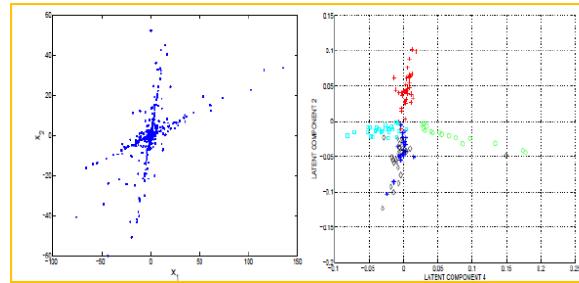
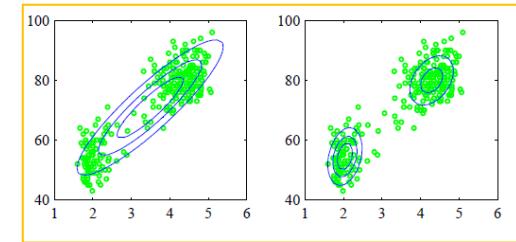
Measurement

Parameters

Analysis by synthesis: generative models

Cautionary note:

Careful distinction between **prior** (say clustering assumption) and **posterior** (how well does the assumption hold...)
ICA: components are independent in prior but not necessarily in the posterior



Supervised Learning in vector spaces

Issues

Selection of model family

How to incorporate unlabeled samples?

Discriminative vs generative learning

A priori knowledge constraints/probabilistic
Learning (ML, MAP, Bayes)

Outlier detection

- Erroneous label
- Unknown label

Model structure optimization

Performance issues

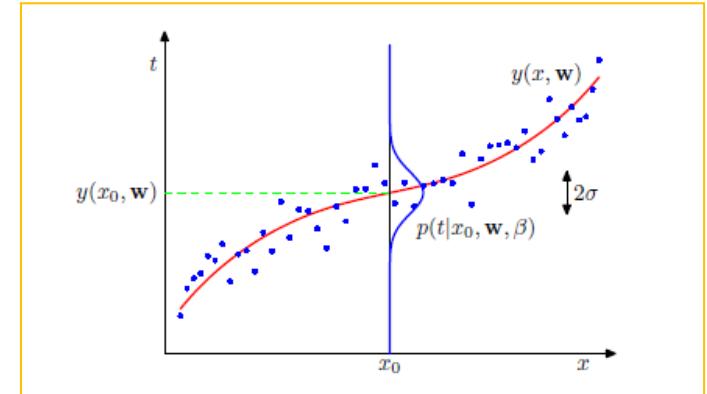
- Generalization
- Confidence
- Consensus methods
- Reproducibility

Visualization/interpretation

- Importance analysis
- Why should I trust you?

$$p(t|\mathbf{x}, \theta)$$

↑
Labels Measurement Parameters



TOWARDS COGNITIVE COMPONENT ANALYSIS

Lars Kai Hansen, Peter Ahrendt, and Jan Larsen

Intelligent Signal Processing,
Informatics and Mathematical Modelling,
Technical University of Denmark B321,
DK-2800 Kgs. Lyngby, Denmark

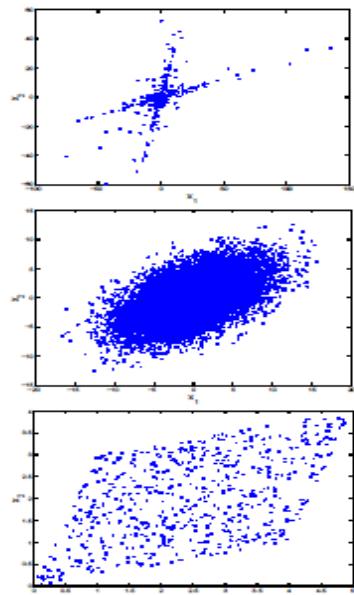


Figure 1. Prototypical feature distributions produced by a linear mixture, based on sparse (top), normal (middle), or dense source signals (bottom), respectively. The characteristic of the sparse signal is that it consists of relatively few large magnitude samples on a background of small signals.

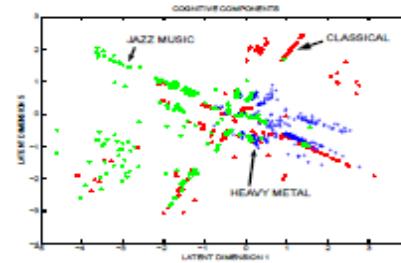


Figure 4. We represent three music tunes (with labels: heavy metal, jazz, classical) by their spectral content in overlapping small time frames ($w = 30\text{msec}$, with an overlap of 10msec , see [22], for details). To make the visualization relatively independent of 'pitch', we use the so-called mel-cepstral representation (MFCC, $K = 13$ coefficients pr. frame). To reduce noise in the visualization we have 'sparsified' the amplitudes. This was achieved simple by keeping coefficients that belonged to the upper 5% magnitude fractile. The total number of frames in the analysis was $F = 10^6$. Latent semantic analysis provided unsupervised subspaces with maximal variance for a given dimension. We show the scatter plot of the data on a 2D subspace within an original 3D PCA. For interpretation we have coded the data points with signatures of the three genres involved: classical (*), heavy metal (diamond), jazz (+). The ICA ray-structure is striking, however, note that the situation is not one-to-one (ray to genre) as in the small test databases. A component (ray) quantifies a characteristic musical 'theme' at the temporal level of a frame (30msec), i.e., an entity similar to the 'phoneme' in speech.

Unsupervised learning

The DNA of music?

Speech: Phonemes, gender, identity

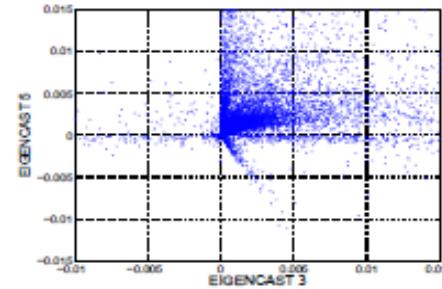


Figure 3. The so-called actor network quantifies the collaborative pattern of 382.000 actors participating in almost 128.000 movies. For visualization we have projected the data onto principal components (LSD) of the actor-actor co-variance matrix. The eigenvectors of this matrix are called 'eigencasts' and they represent characteristic communities of actors that tend to co-appear in movies. The network is extremely sparse, so the most prominent variance components are related to near-disjunct sub-communities of actors with many common movies. However, a close up of the coupling between two latent semantic components (the region $\sim (0,0)$) reveals the ubiquitous signature of a sparse linear mixture: A pronounced 'ray' structure emanating from $(0,0)$. We speculate that the cognitive machinery developed for handling of independent events can also be used to locate independent sub-communities, hence, navigate complex social networks, a hallmark of successful cognition.





Supervised learning of sentiment

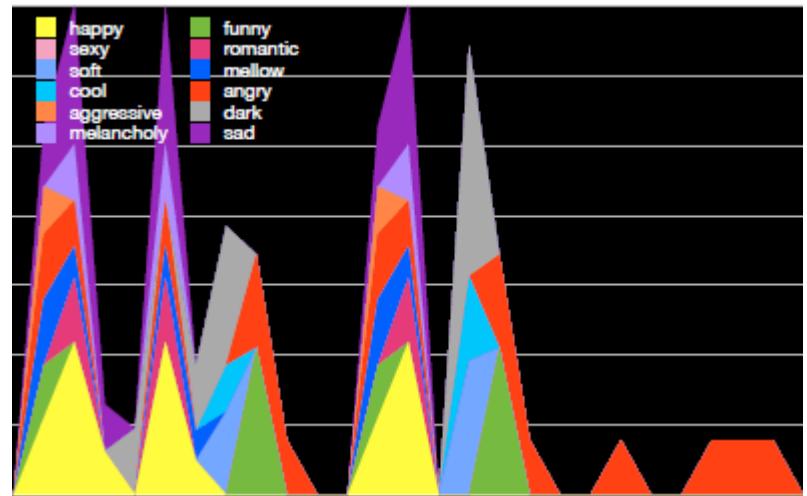
A step towards understanding subjectivity, opinion

Important to many services

Recommender (Amazon reviews)

Information navigation, e.g. navigating music

Oasis "Wonderwall"
Emotional content in the
song lyrics through time



M.K. Petersen: Modeling media as latent semantics
based on cognitive components (Ph.D. Thesis, DTU, 2010)

Business applications of non-linear signal processing

- International
 - Google, Yahoo!, Microsoft
 - Recommender services: Amazon, NetFlix, Spotify
- DK
 - Hearing aids: Oticon, Widex, GN
 - Communication, digital media: Adform,
 - Monitoring: MAN B&W Diesel, Brüel & Kjær

Introducing Amazon Virtual Private Cloud

Securely bridge your IT infrastructure to the AWS cloud.

[Learn More...](#)



Learning objectives

A student who has met the objectives of the course will:

Be able to discuss fundamental concepts in machine learning: Generalizability, likelihood functions, Bayesian modeling, and the bias-variance trade-off.

Be able to discuss fundamental concepts of signal detection: Bayesian decision theory, the role of posterior probabilities and loss functions.

Establish likelihood functions for signal modeling and detection.

Design adaptive linear and non-linear signal modeling and detection systems

Use cross validation to obtain un-biased performance estimates.

Evaluate adaptive linear and non-linear signal modeling and detection systems

Explain implementations of adaptive systems and evaluation methods in Matlab.

Be familiar with applications of machine learning in audio signal detection.

Be familiar with applications of machine learning in bio-medical data.

Give a verbal presentation of results obtained in the course's Matlab exercises

What kinds of math are necessary?

Vectors and matrices

Eigenvectors and eigenvalues

Trace and determinant

Sequences and summation

Taylor and Fourier series

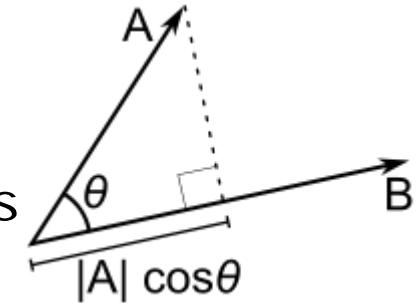
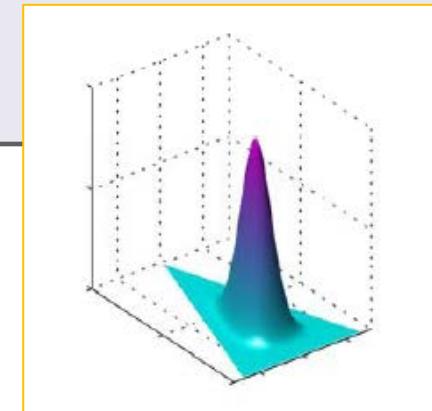
Functions on vector spaces and their derivatives
(gradient and Hessian)

Integrals in one and “many” dimensions

Optimization theory:

Gradient descent,

Newton methods



slice

$$\begin{bmatrix} 2 & 1 & 2 \\ 3 & 2 & 3 \\ 4 & 1 & 1 \end{bmatrix} \times \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 2*1 + 1*0 + 2*1 \\ \quad \\ \quad \end{bmatrix}$$

Probabilities

We define the probability of an event as the limit of a frequency in a sampling process.

Flip a coin, count the number of heads (N_{head})

$$P(\text{head}) = \lim_{N \rightarrow \infty} \frac{N_{\text{head}}}{N}$$



Joint probabilities



Probabilities are limits of relative # counts (frequencies)

Event characterized by two “values” (x, y)

We observe pairs (x, y)

Examples

- Signal level vs bit (0,1)
- Foot length vs shoe size
- Image of surface vs quality of object

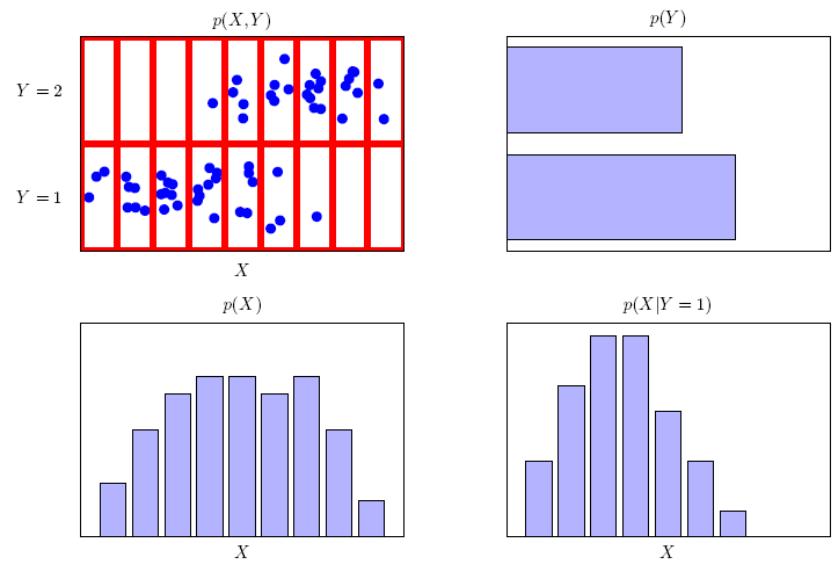


Figure 1.11 An illustration of a distribution over two variables, X , which takes 9 possible values, and Y , which takes two possible values. The top left figure shows a sample of 60 points drawn from a joint probability distribution over these variables. The remaining figures show histogram estimates of the marginal distributions $p(X)$ and $p(Y)$, as well as the conditional distribution $p(X|Y = 1)$ corresponding to the bottom row in the top left figure.

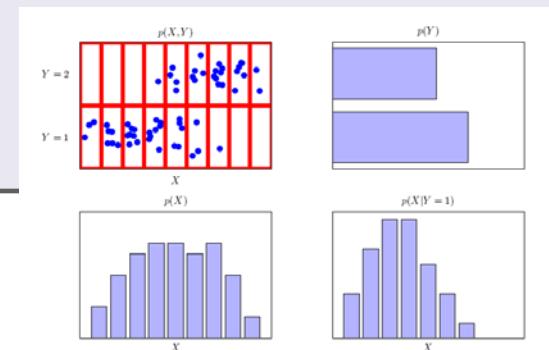


Figure 1.11 An illustration of a distribution over two variables, X , which takes 9 possible values, and Y , which takes two possible values. The top left figure shows a sample of 60 points drawn from a joint probability distribution over these variables. The remaining figures show histogram estimates of the marginal distributions $p(X)$ and $p(Y)$, as well as the conditional distribution $p(X|Y = 1)$ corresponding to the bottom row in the top left figure.

Conditional probability

- Bayes' theorem

$$p(x, y) = \lim_{N \rightarrow \infty} \frac{N_{x,y}}{N} = \lim_{N \rightarrow \infty} \frac{\frac{N_{x,y}}{N_x}}{\frac{N_x}{N}} = p(y | x)p(x)$$

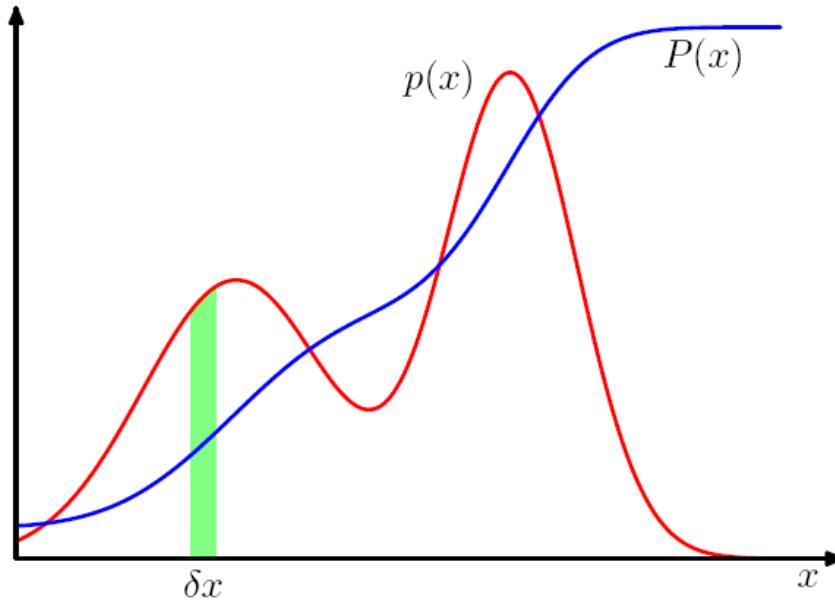
$$p(x, y) = \lim_{N \rightarrow \infty} \frac{N_{x,y}}{N} = \lim_{N \rightarrow \infty} \frac{N_{x,y}}{N_y} \frac{N_y}{N} = p(x | y)p(y)$$

Probability density functions

$$P(x \in [a, b]) = \int_a^b p(x)dx$$

- Events with finite probability are intervals

Figure 1.12 The concept of probability for discrete variables can be extended to that of a probability density $p(x)$ over a continuous variable x and is such that the probability of x lying in the interval $(x, x + \delta x)$ is given by $p(x)\delta x$ for $\delta x \rightarrow 0$. The probability density can be expressed as the derivative of a cumulative distribution function $P(x)$.



Expectations computed from the pdf

$$\mathcal{E}(f(x)) = \int_{\text{Domain of } x} f(x)p(x)dx$$

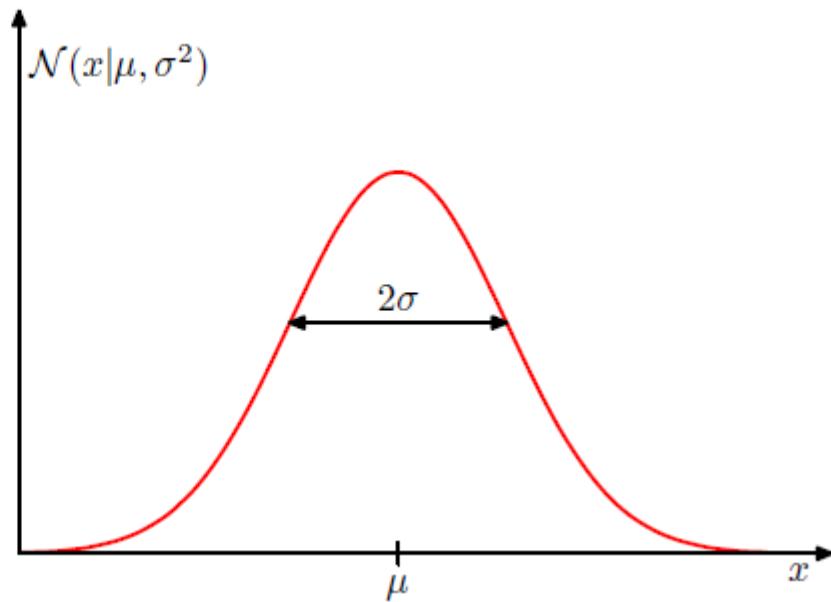
$$P(x \in \text{Domain of } x) = \int_{\text{Domain of } x} p(x)dx = 1$$

$$\mathcal{E}(x) \equiv \mu = \int_{\text{Domain of } x} xp(x)dx$$

The typical "spread" of the data

$$\sigma = \sqrt{\int_{\text{Domain of } x} (x - \mu)^2 p(x) dx}$$

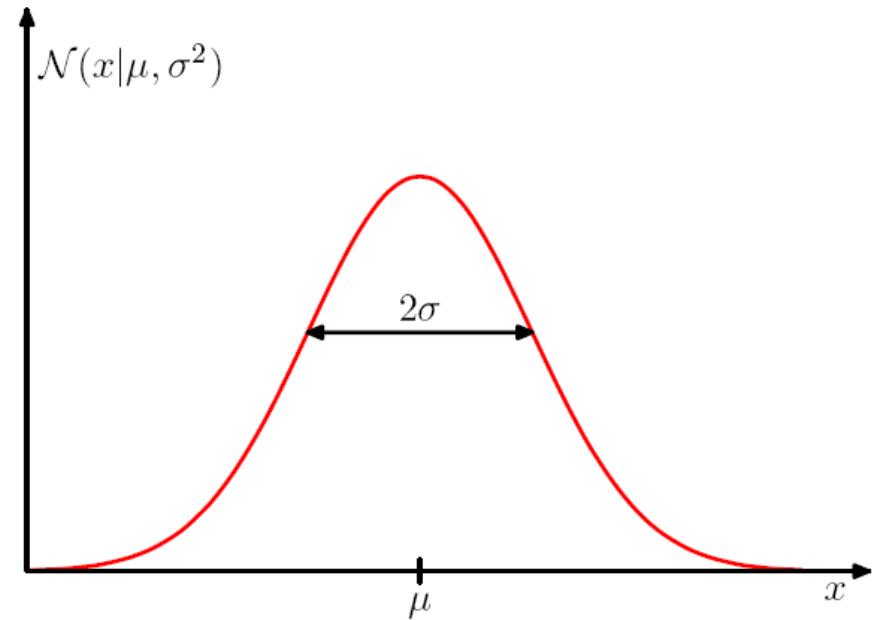
Figure 1.13 Plot of the univariate Gaussian showing the mean μ and the standard deviation σ .



The normal distribution as a signal model

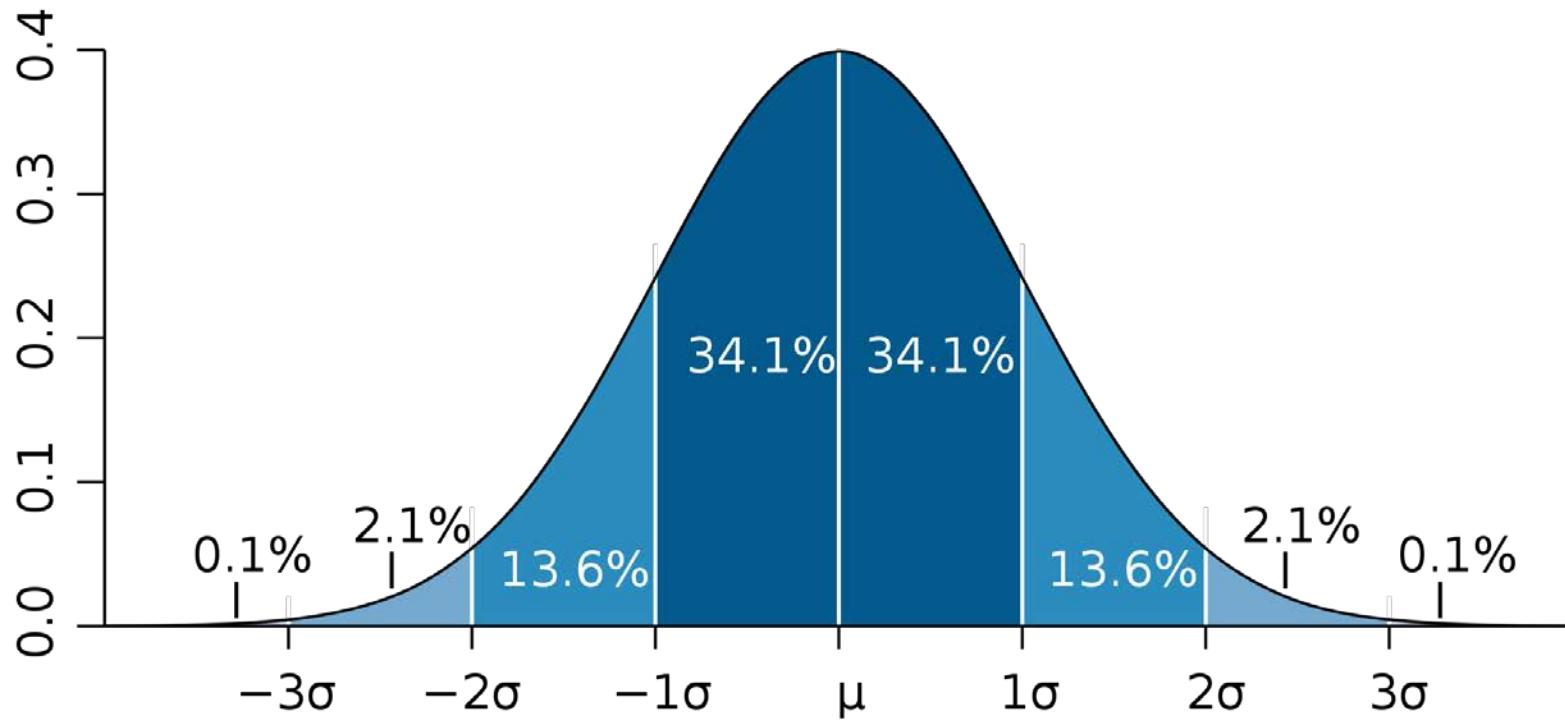
Figure 1.13 Plot of the univariate Gaussian showing the mean μ and the standard deviation σ .

If events can be described as a “mean effect” with additive noise



$$p(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x - \mu)^2\right)$$

Probabilities under the normal pdf curve



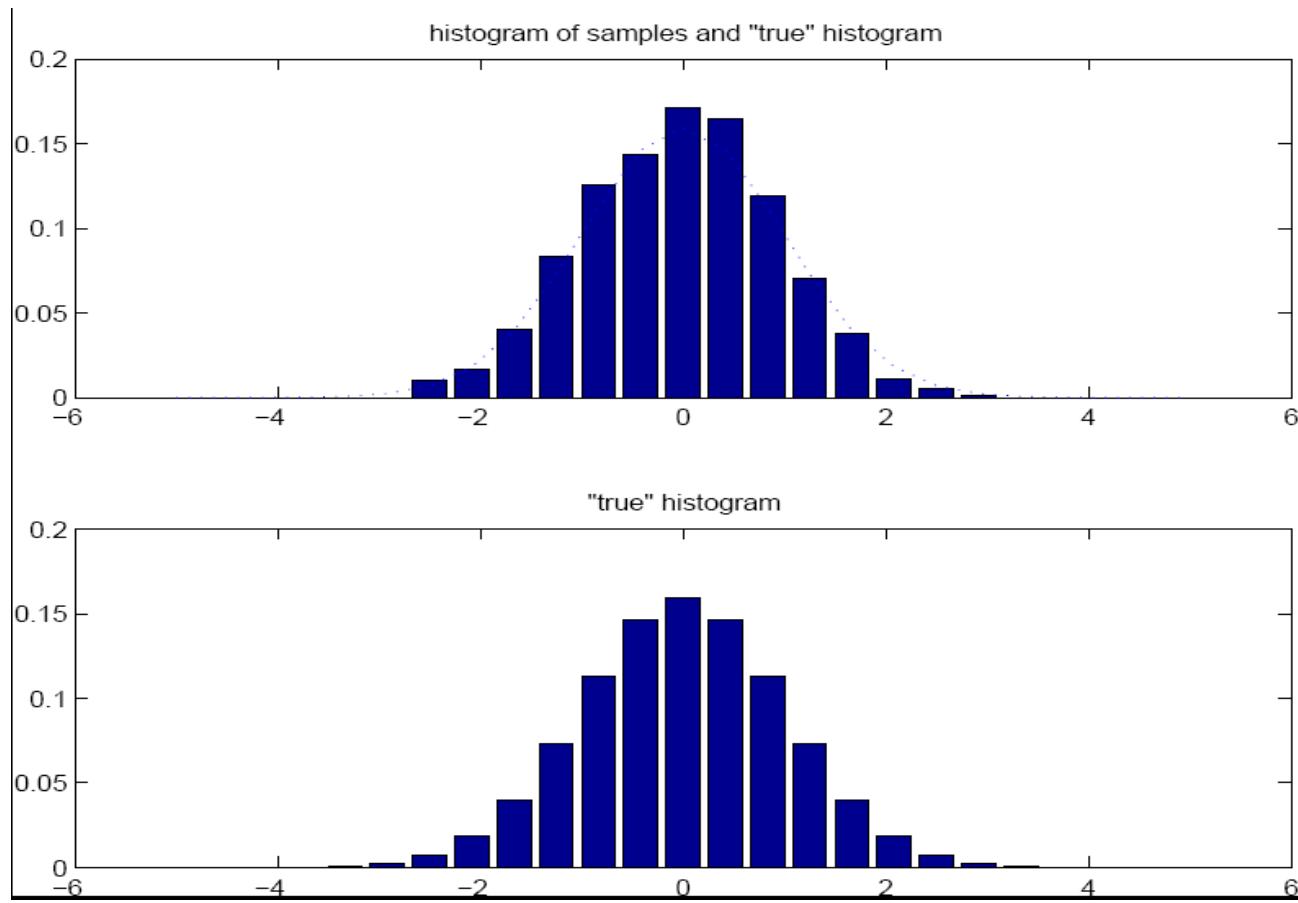
Parameters in the normal distribution

$$p(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x - \mu)^2\right)$$

$$\mu = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x - \mu)^2\right) dx$$

$$\sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x - \mu)^2\right) dx$$

Models vs data



$$\mu = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x - \mu)^2\right) dx$$

$$\mu_{\text{ML}} = \frac{1}{N} \sum_{n=1}^N x_n$$

Signal detection problem

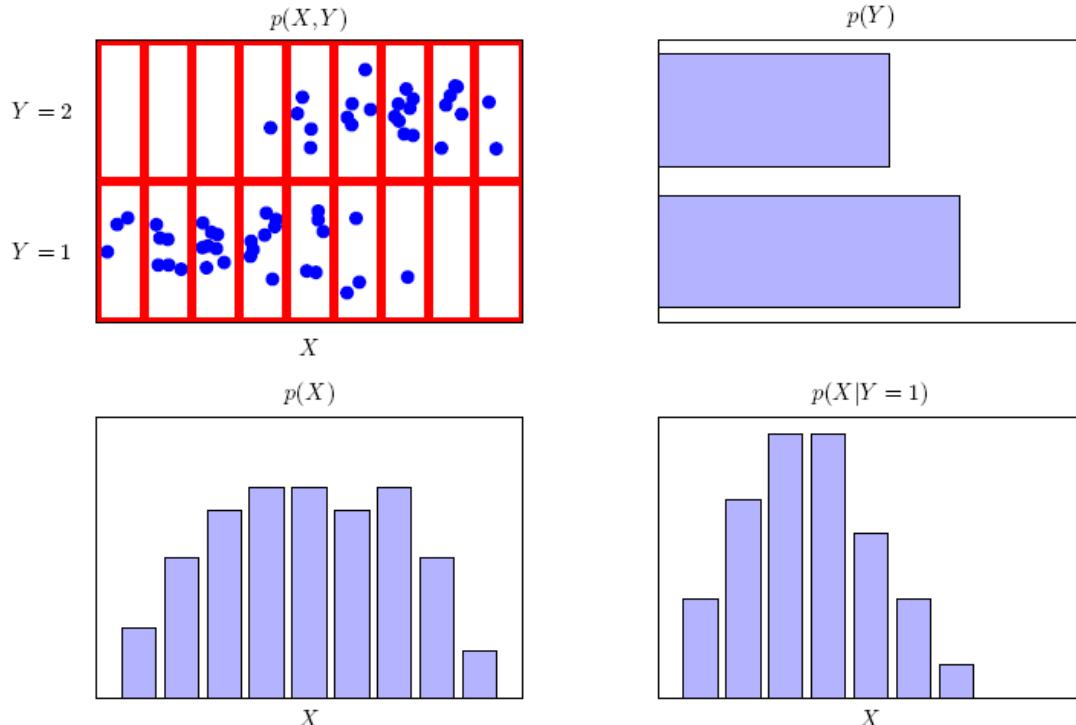


Figure 1.11 An illustration of a distribution over two variables, X , which takes 9 possible values, and Y , which takes two possible values. The top left figure shows a sample of 60 points drawn from a joint probability distribution over these variables. The remaining figures show histogram estimates of the marginal distributions $p(X)$ and $p(Y)$, as well as the conditional distribution $p(X|Y = 1)$ corresponding to the bottom row in the top left figure.

$$P(\mathcal{C}_k, \mathbf{x}) = p(\mathbf{x}|\mathcal{C}_k)P(\mathcal{C}_k)$$

$$P(\mathcal{C}_k, \mathbf{x}) = P(\mathcal{C}_k|\mathbf{x})p(\mathbf{x})$$

$$P(\mathcal{C}_k|\mathbf{x}) = \frac{p(\mathbf{x}|\mathcal{C}_k)P(\mathcal{C}_k)}{p(\mathbf{x})}$$

$$p(\mathbf{x}|\mathcal{C}_k) = \frac{P(\mathcal{C}_k|X^l)p(\mathbf{x})}{P(\mathcal{C}_k)}$$

$$\sum_{k=1}^c P(\mathcal{C}_k|\mathbf{x}) = 1$$

$$\sum_{k=1}^c p(\mathbf{x}|\mathcal{C}_k)P(\mathcal{C}_k) = p(\mathbf{x})$$

Signal detection: Bayes decision theory

- A signal detection system (or pattern classifier) provides a rule for assigning a measurement to a given signal category (class)
- Hence, a classifier divides measurement space (feature space) into disjoint regions $\mathcal{R}_1, \mathcal{R}_2, \dots, \mathcal{R}_c$, such that measurements that fall into region \mathcal{R}_k are assigned with class \mathcal{C}_k .
- Boundaries between regions are denoted decision surfaces or decision boundaries

Signal Detection: Bayes decision theory

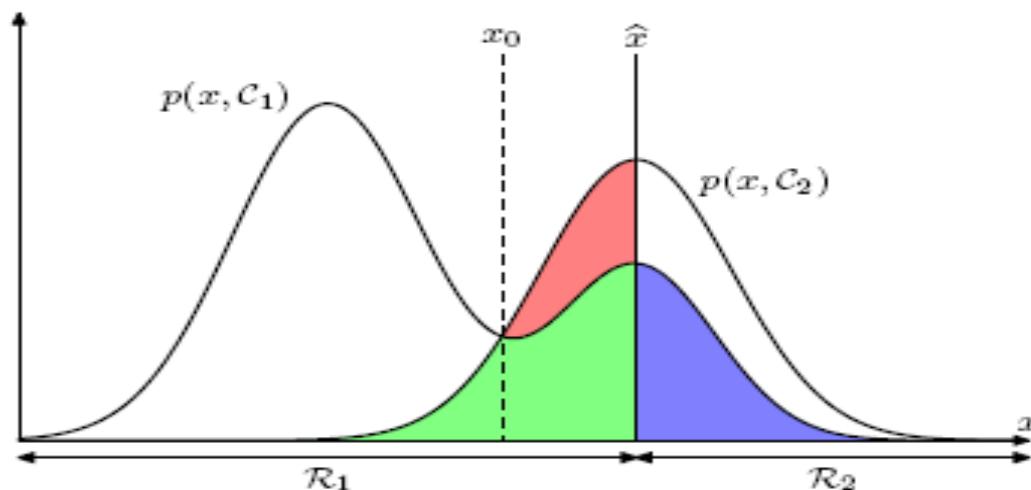


Figure 2: Schematic plot of the densities for a measured signal drawn from either of two populations $\mathcal{C}_1, \mathcal{C}_2$

$$\begin{aligned} P(\text{error}) &= P(x \in \mathcal{R}_2, \mathcal{C}_1) + P(x \in \mathcal{R}_1, \mathcal{C}_2) \\ &= P(x \in \mathcal{R}_2 | \mathcal{C}_1)P(\mathcal{C}_1) + P(x \in \mathcal{R}_1 | \mathcal{C}_2)P(\mathcal{C}_2) \\ &= \left(\int_{\mathcal{R}_2} p(x | \mathcal{C}_1) dx \right) P(\mathcal{C}_1) + \left(\int_{\mathcal{R}_1} p(x | \mathcal{C}_2) dx \right) P(\mathcal{C}_2) \end{aligned}$$

- The probability of error is minimized if we assign points to \mathcal{R}_1 , whenever $p(x | \mathcal{C}_1)P(\mathcal{C}_1) > p(x | \mathcal{C}_2)P(\mathcal{C}_2)$

Signal detection: Use posterior for decision

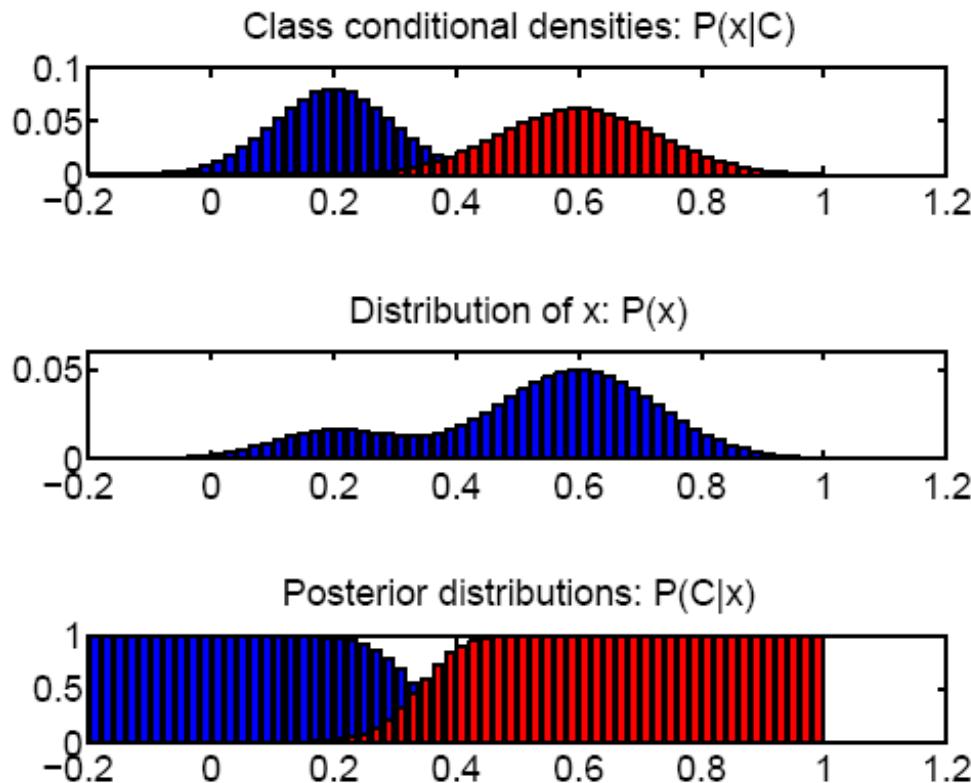


Figure 1: Schematic plot of the histograms for a measured signal drawn from either of the populations $\mathcal{C}_1, \mathcal{C}_2$, density of x , and the corresponding posteriors $P(\mathcal{C}|X)$'s

$$P(\mathcal{C}_k, \mathbf{x}) = p(\mathbf{x}|\mathcal{C}_k)P(\mathcal{C}_k)$$
$$P(\mathcal{C}_k, \mathbf{x}) = P(\mathcal{C}_k|\mathbf{x})p(\mathbf{x})$$

$$P(\mathcal{C}_k|\mathbf{x}) = \frac{p(\mathbf{x}|\mathcal{C}_k)P(\mathcal{C}_k)}{p(\mathbf{x})}$$
$$p(\mathbf{x}|\mathcal{C}_k) = \frac{P(\mathcal{C}_k|X^l)p(\mathbf{x})}{P(\mathcal{C}_k)}$$

$$\sum_{k=1}^c P(\mathcal{C}_k|\mathbf{x}) = 1$$
$$\sum_{k=1}^c p(\mathbf{x}|\mathcal{C}_k)P(\mathcal{C}_k) = p(\mathbf{x})$$

Classification with asymmetric loss

$$\begin{aligned} R_k &= \sum_{j=1}^c L_{k,j} \int_{\mathcal{R}_j} p(\mathbf{x}|\mathcal{C}_k) d\mathbf{x} \\ R &= \sum_{k=1}^c R_k P(\mathcal{C}_k) \\ &= \sum_{j=1}^c \int_{\mathcal{R}_j} \sum_{k=1}^c L_{k,j} p(\mathbf{x}|\mathcal{C}_k) P(\mathcal{C}_k) d\mathbf{x} \end{aligned}$$

$$\sum_{k=1}^c L_{k,j} p(\mathbf{x}|\mathcal{C}_k) P(\mathcal{C}_k) < \sum_{k=1}^c L_{k,i} p(\mathbf{x}|\mathcal{C}_k) P(\mathcal{C}_k)$$