Neuroinformatics Lecture (L1)

Prof. Dr. Gordon Pipa
Institute of Cognitive Science, University of Osnabrück

My Short CV

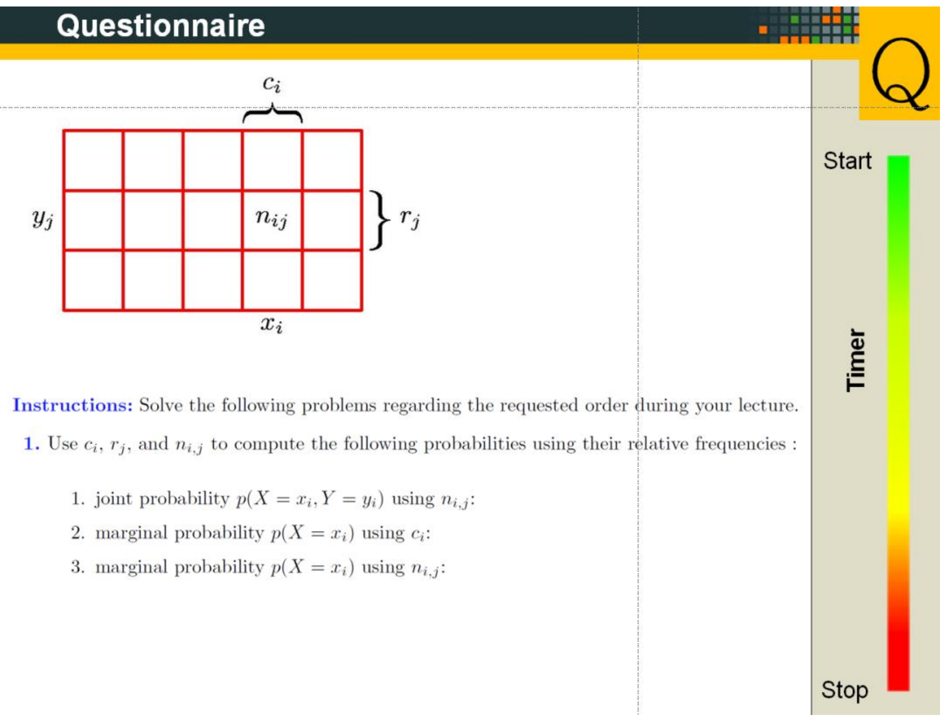
- **Since 2011**, professor and chair of the Neuroinformatics Department at the *Institute of Cognitive Science* in Osnabrück
- **2010: Habilitation in Biology** at the TU-Darmstadt
- **Till 2010**, research group leader at *Max-Planck Institute for Brain Research in Frankfurt* and the *Frankfurt Institute of Advanced Studies*
- **2007–2009**: Research fellow at the *Massachusetts Institute for Technology* and *Harvard MGH*
- **2006**: Visiting Professor at the *Balearic Institute of Complex Sciences*
- **2002–2006: PhD thesis in informatics** at *Max-Planck Institute for Brain Research in Frankfurt* and *TU-Berlin*
- **2002: Diploma in theoretical physics** with a focus on stochastic and correlated systems

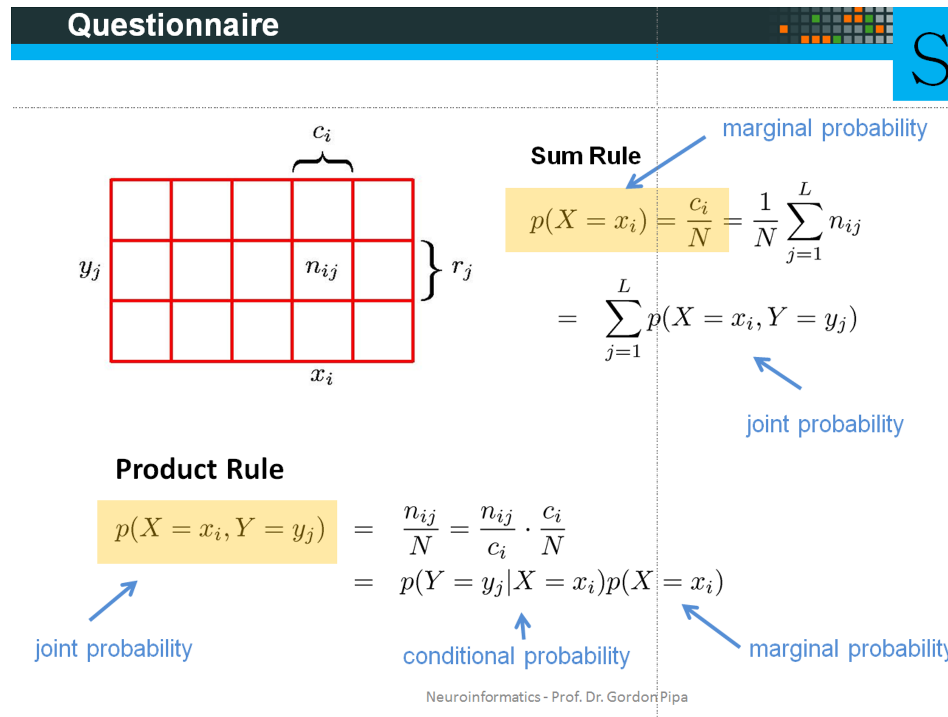
- Lectures are available via StudIP as Flash movies after the lecture was given.
- Exercise sheets are available via StudIP as a PDF.
- Solutions for the exercises will be available via StudIP as a PDF after the due date.
- Questionnaires and their solutions will be available via StudIP as a PDF after the lecture.

Color codes

Orange Q:
Questionnaire Questions

Blue S:
Questionnaire Solutions





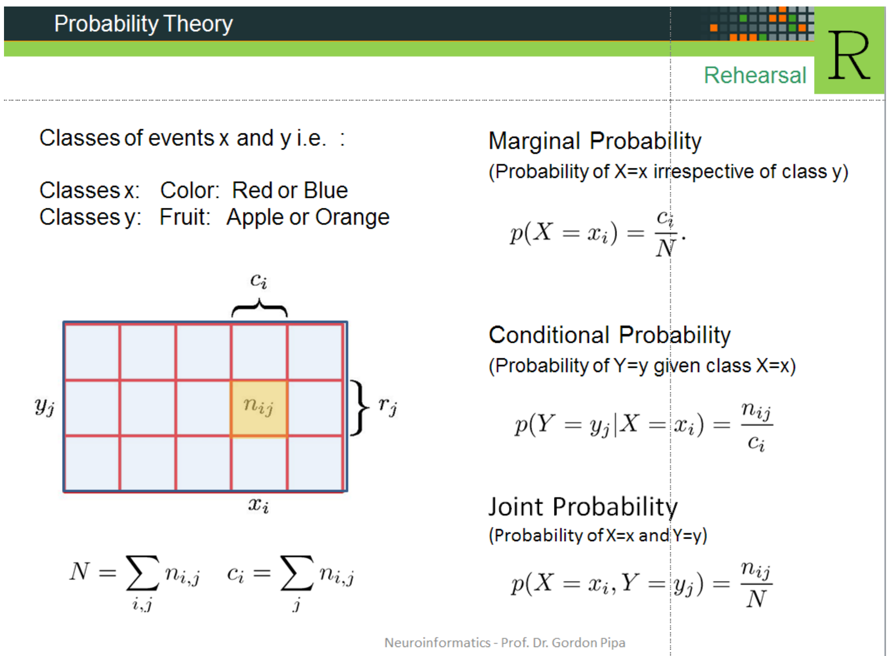
Color codes

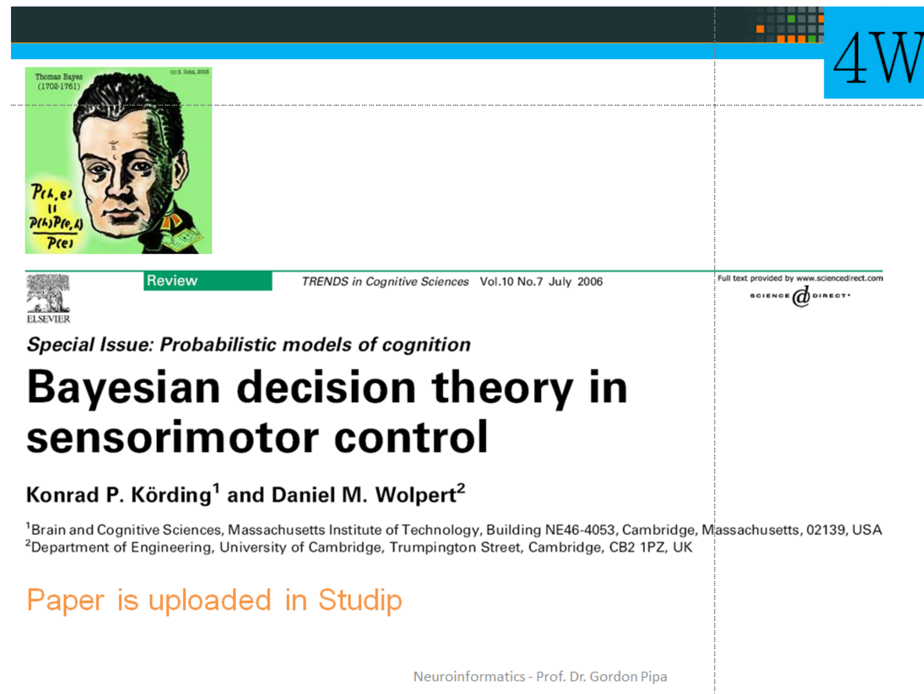
Blue 4W:

- **W**hat is this it all about ?
- **W**hat can I use it for ?
- **W**hy should I learn it ?
- Potential BsC and MsC thesis topics

Green R:

Rehearsal of last lecture



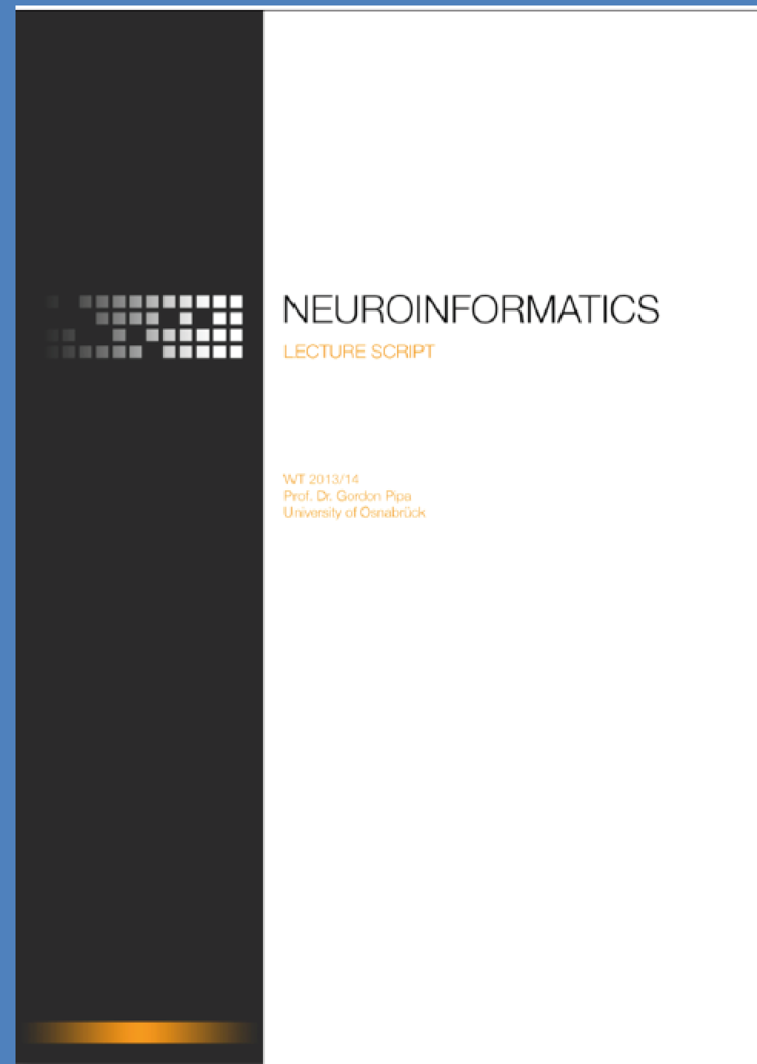


Importance of the lecture

- Participate !
- Use the lecture to ask question !
- Working from home does not replace the lecture.
- The lecture is important since it allows to US to interact. This gives you clues about the relevant topics for the exam !
- In the past we saw a strong correlation between grades in the exam and active presence in the lecture
- If you have a questions: Stop me and wave you hands, … .

Lecture Script

- Comes in a ZIP with Matlab files included
- Always work in progress !
- Migth be updated several times throughout the semester
- Version number will be indicated in file names
- If you find mistakes or if something is incomprehensible, write an email to niscript@ni.uos.de



Content Lecture Script Version 5.b

Contents

1 Recommended Literature 4

2 Introduction 7
- 2.1 Why do we need statistical modeling? 7
- 2.2 Example: Model fitting 7
- 2.3 How is the model constructed? 7
- 2.4 Example: Linear combinations of polynomials 8
- 2.5 Outlook: What is it useful for? 8

3 Probability theory and related topics 10
- 3.1 Marginal probability 10
- 3.2 Conditional probability 11
- 3.3 Joint probability 11
- 3.4 Sum and product rule 11
- 3.5 Bayes' theorem 12
- 3.6 Summary: Probability theory and Bayes 13
- 3.7 Continuous random variables probability density functions 13
- 3.8 Transformation of densities 15
  - 3.8.1 Example: Transformation of densities 16
- 3.9 Expectations 17

4 Graphical models 19
- 4.1 Bayesian networks 19
- 4.2 Statistical independence of variables 21
  - 4.2.1 Conditional independence of variables 21
- 4.3 D-separation: Statistical independence in graphs 21
  - 4.3.1 Head-to-tail nodes 21
  - 4.3.2 Fixed head-to-tail nodes 23
- 4.4 Head-to-head nodes 23
- 4.5 Fixed head-to-head nodes 24
- 4.6 Tail-to-tail nodes 24
  - 4.6.1 Fixed tail-to-tail nodes 25
- 4.7 Overview & more 25

5 Probability distributions 27
- 5.1 Discrete distributions 27
  - 5.1.1 Bernoulli distribution 27
  - 5.1.2 Binomial distribution 28
  - 5.1.3 Poisson distribution 28
- 5.2 Continuous distributions 29
  - 5.2.1 Normal distribution 29
  - 5.2.2 Gamma distribution 31
  - 5.2.3 Beta distribution 32
  - 5.2.4 χ^2 distribution 33
- 5.3 Summary 35

6 Model Fitting via Maximum Likelihood 36
- 6.1 The Maximum Likelihood Approach - Univariate Data Sets 36
  - 6.1.1 Rationale of The ML Approach 37
  - 6.1.2 Setting up the likelihood function (for a single data point) 37

Neuroinformatics lecture script

  - 6.1.3 Joint Likelihood (multiple data points) 38
  - 6.1.4 Deriving the ML Estimators 38
  - 6.1.5 Normal Distributions: Deriving μ_{ML} 38
  - 6.1.6 Normal Distributions: σ^2_{ML} and bias 40
  - 6.1.7 Overview 41
- 6.2 Bivariate data: The linear regression 42
  - 6.2.1 Transition from univariate data 42
  - 6.2.2 The underlying model 43
  - 6.2.3 Setting up our model 43
  - 6.2.4 Setting up the likelihood function (bivariate) 44
  - 6.2.5 Maximizing the likelihood: The method of least squares 44
  - 6.2.6 The Moore-Penrose-pseudoinverse 46
- 6.3 Examples of model fitting (bivariate) 46
  - 6.3.1 The pen and paper method 46
  - 6.3.2 Fitting a model to data using Matlab 48

7 Basis functions 51
- 7.1 Polynomial basis functions 51
- 7.2 Gaussian basis functions 51
- 7.3 Sigmoidal basis functions 52
- 7.4 Sinusoid basis functions 52
- 7.5 Bin-based basis functions 54
  - 7.5.1 Continuity in piecewise polynomials 54
  - 7.5.2 B-splines 55

8 Generalizability, validation and model selection 57
- 8.1 Training- and validation data 57
- 8.2 Number of basis functions, over- and underfitting 57
- 8.3 Cross-validation 58

9 Model selection: Nested models, likelihood ratio and deviance 61
- 9.1 Nested and saturated models 61
- 9.2 Likelihood ratios 61
- 9.3 Saturated models 63

10 Bias-variance-decomposition 64

11 Regularization 66

12 Bayesian regression 70
- 12.1 Conjugate prior and MAP estimates for Bernoulli 71
- 12.2 Sequential updates to the posterior/ prior 72
- 12.3 Choosing the right prior 73
- 12.4 Bayesian regression and regularization 73

13 Model selection: Akaike Information Criterion 75
- 13.1 The Kullback-Leibler divergence 75
- 13.2 AIC: A bias-corrected approximation of the KL divergence 77

14 Model selection: Bootstrap and EIC 78

Neuroinformatics lecture script

2 Introduction

2.1 Why do we need statistical modeling?

When collecting data in the real world - e.g. recording the spiking of a neuron - the data recorded is usually subject to random fluctuations (noise). Often the noise is intense enough to nearly completely hide the true structure of the data. We use statistical modeling to reconstruct this underlying structure from noisy data.

2.2 Example: Model fitting

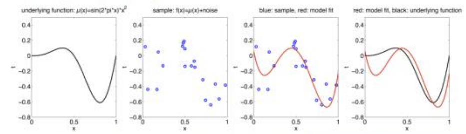


Figure: An example of statistical modeling. In the first plot, we see the original function (also called underlying function or underlying model), which represents some course of a variable. For instance, this could be voltage over time, maybe a neuron's membrane potential, but also EEG- or behavioral data. The underlying function is unknown to us in the real world. Here, all we ever have is measured data, which is subject to noise, i.e. the observations are a superposition of the original (or underlying) function and noise (second plot). In order to uncover the structure of the underlying function, we fit a flexible function to the observed data, the result is what we call our model (third plot, red line). You can see a comparison of the estimated model (red) with the underlying model (black) in the fourth plot. Bear in mind that this comparison can't be done in the real world, since we never know the exact underlying function (black).

For a better understanding, the Matlab code behind many of the figures in this script is provided with the script. Often, an algorithmic approach is easier to understand than an analytic one. Thus, if there is anything you don't get, and there's a related figure, reading the code and playing around with it is highly recommended. We aimed to keep the code simple and structured, so basic programming skills should enable you to understand the code in most cases. The Matlab file for this figure is provided in *ch2fig1.m*. In this case, you probably won't understand what the code is doing just yet. However, you might want to come back to it later.

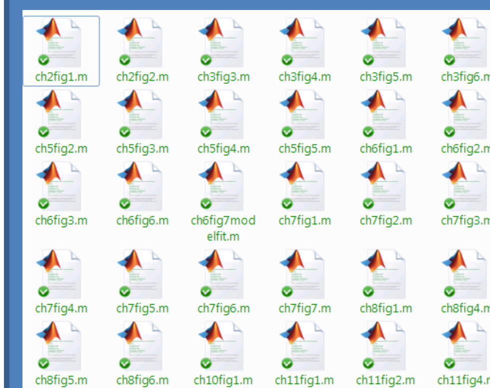
2.3 How is the model constructed?

Our estimated model (e.g. the red curve above) is a weighted linear combination of functions - so called basis functions - and consequently a function itself. The basis functions within a model usually stem from the same family of functions, e.g. polynomials, gaussians or cubic splines. In the first part of the lecture (and the script), we will mostly deal with polynomials as basis functions. Accordingly, our n basis functions are $x^0, x^1, x^2, ..., x^n$. In principle, we are free to choose the number of basis functions ourselves. The corresponding weights are called $w_0, w_1, w_2, ..., w_n$. Thus, our polynomial model is defined as

$$y = w_0 + w_1 x + w_2 x^2 + ... + w_n x^n$$

or in a different notation

$$y(\vec{w}, x) = \sum_i (w_i \cdot x^i).$$



Matlab files for reproducing figures in the script

Structure of the Lecture series:

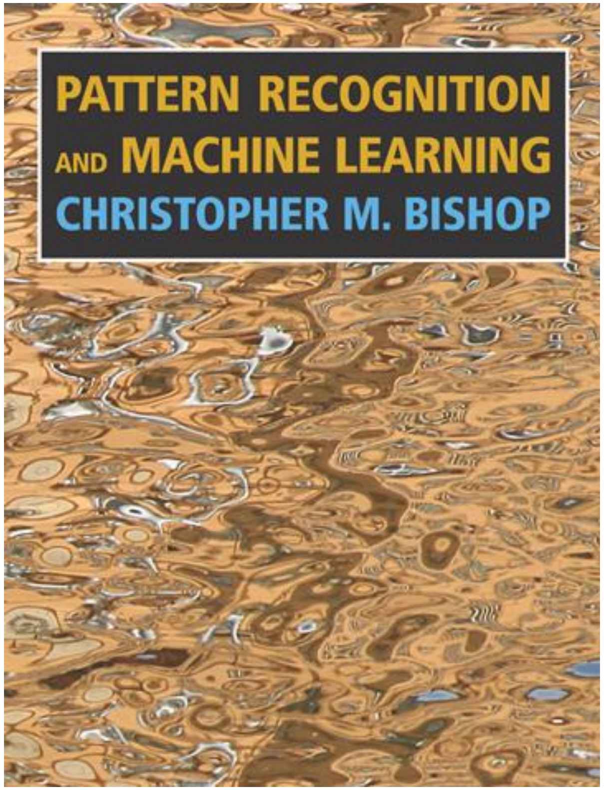
L1-8 : Probability and Maximum Likelihood
L9-10 : Regression Problem
L11 : Brain computer Interface
L12 : Nested Models
L13 : Families of Basis Functions
L14 : Bias Variance Decomposition
L15 : Modell selction AIC
L16-17 : Bootstrap EIC
L18 : Deep Brain stimulation (Theory and Practice)
L19 : Non-Gaussian Regression
L20 : Christmas Lecture (Awakenings / Zeit des Erwachens)
L21 : Rehearsal of what we learned so far
L22 : Regularization
L23 : Bayes Regression – Maximum a posteriori
L24-25 : Classification
L26-27 : Expectation Maximization
L28 : QA

Written EXAM : 8th feb. from 4-8 p.m.

Recommended Books

- **Pattern Recognition and Machine Learning**
  by Christopher M. Bishop, Publisher: Springer
  http://research.microsoft.com/en-us/um/people/cmbishop/prml/

  Very nice book, from easy to advanced



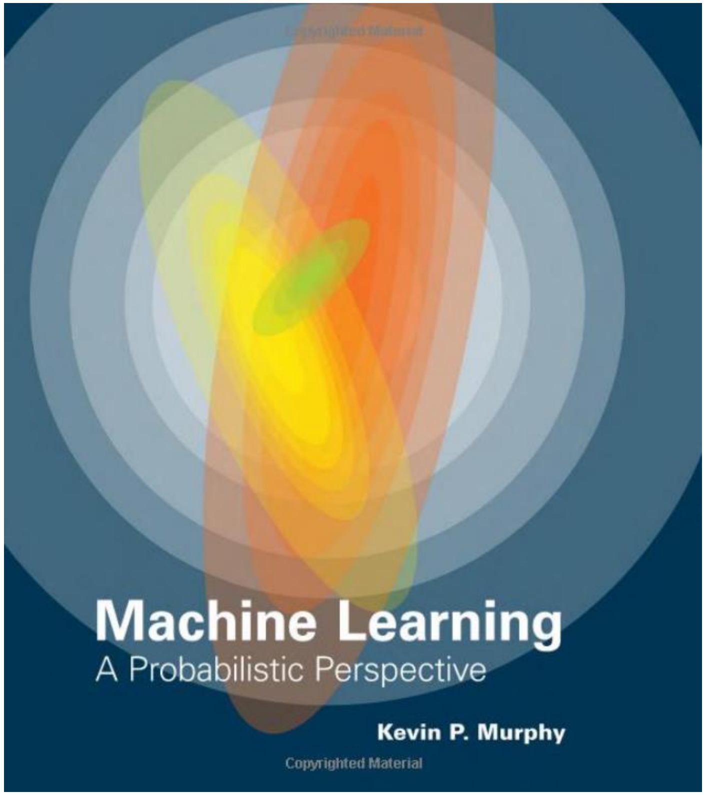
Recommended Books

Machine Learning: A Probabilistic Perspective

by Kevin Murphy, The MIT Press (7. September 2012)

Very nice book, from easy to advanced. Gives nice inside in BIG-Data analysis and what *Google* does !

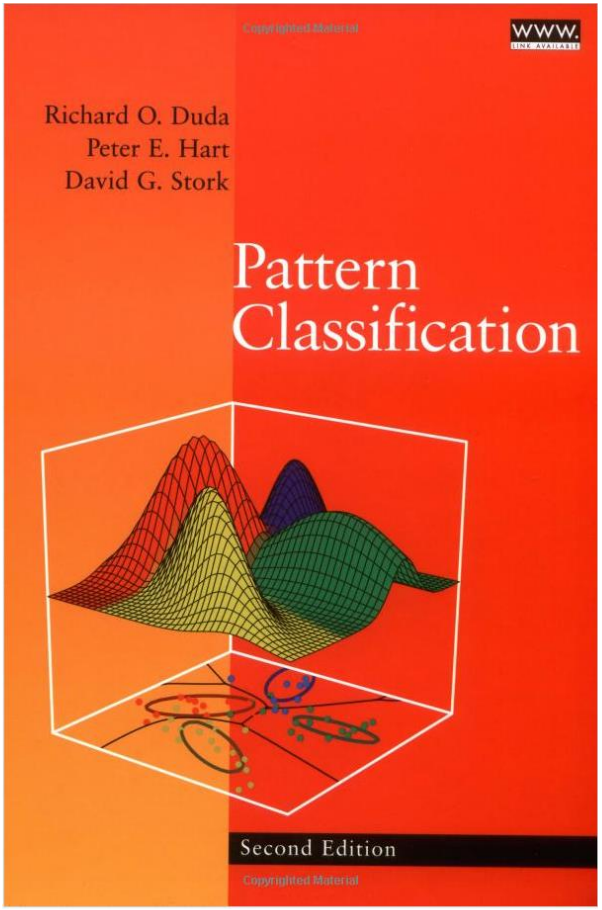
http://www.cs.ubc.ca/~murphyk/MLbook/



Recommended Books

- **Pattern Classification (2nd Edition)**
  (Student Solutions Manual also available)
  by Duda, Hart, and Storck

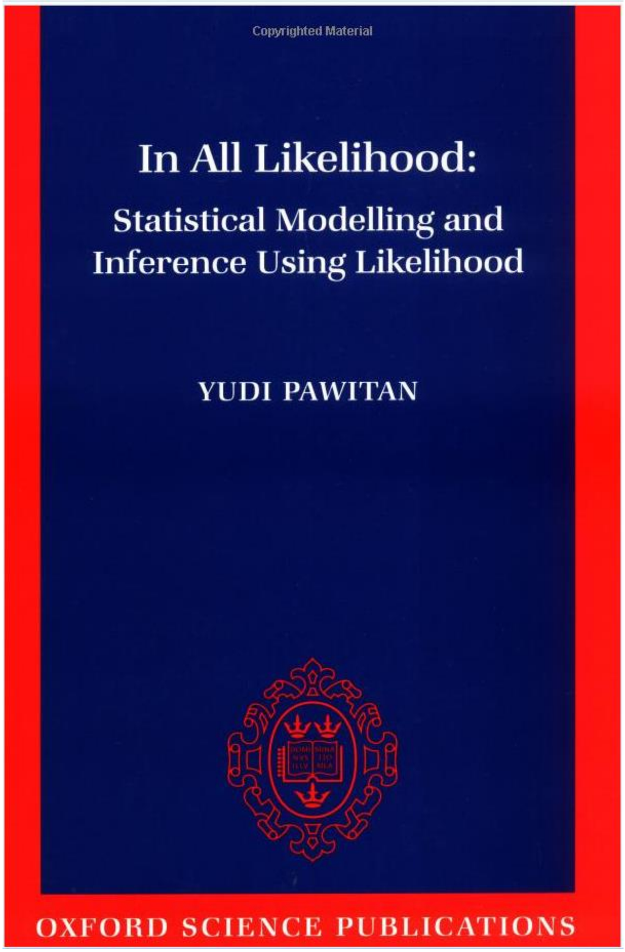
  Emphasizes classification, from easy to advanced



Recommended Books

- **In All Likelihood**
  By Y. Pawitan, Publisher: Oxford University Press

  Rather Advanced

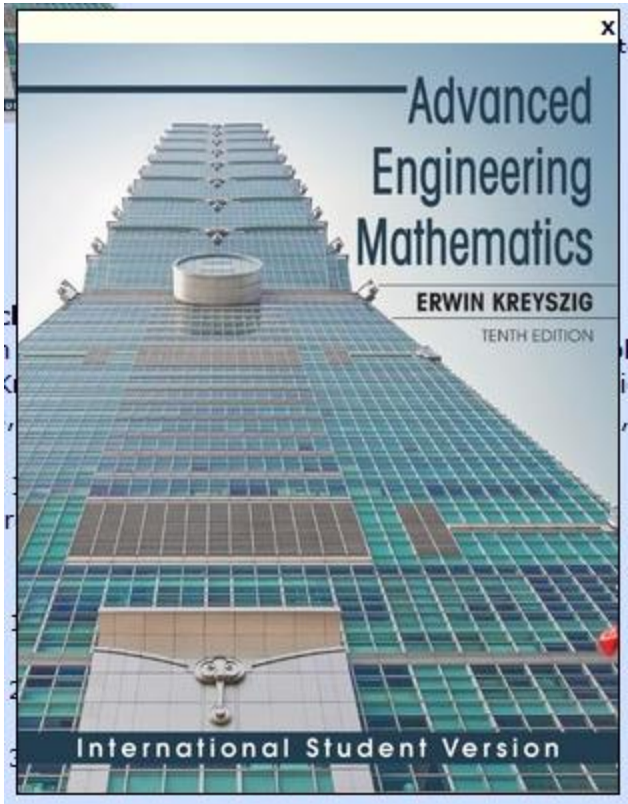


Recommended Books

- **Advanced Engineering Mathematics Textbook (Student Solutions Manual also available)**
  by Erwin Kreyszig
  Mai 2011
  61,90 Euro

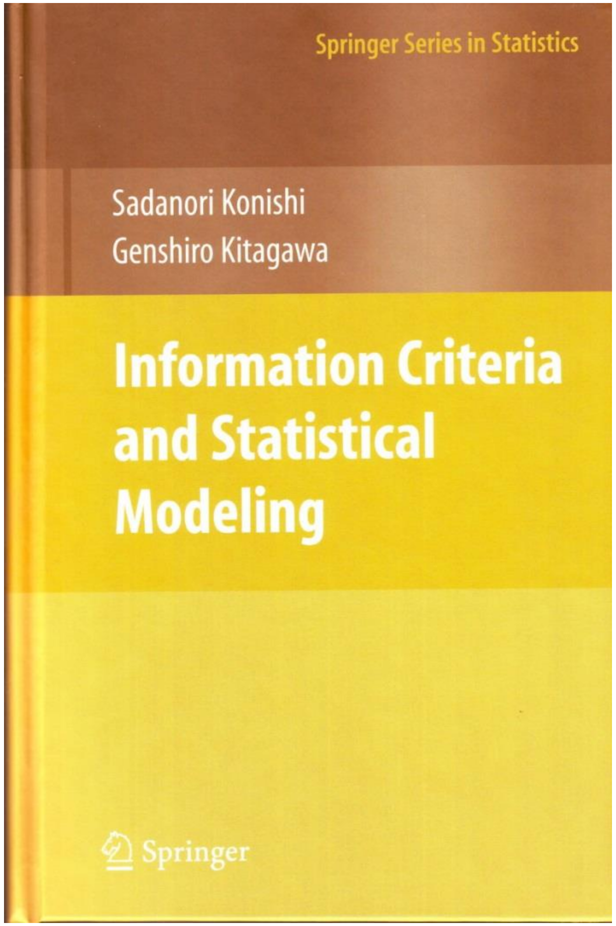
  2011. 1152 pages, Softcover
  ISBN-10: 0-470-64613-6
  ISBN-13: 978-0-470-64613-7 - John Wiley & Sons

  Great Math book. Gives examples and combines these with rigorous proofs.



Recommended Books

Edition 2007
273 pages,
Hard cover
English
Springer, Berlin
ISBN-10: 0387718869
ISBN-13: 9780387718866



Advanced!
Great book for people who want to know all about model selection

Neuroinformatics Lecture

Motivation for this course

Motivation: What is Neuroinformatics about

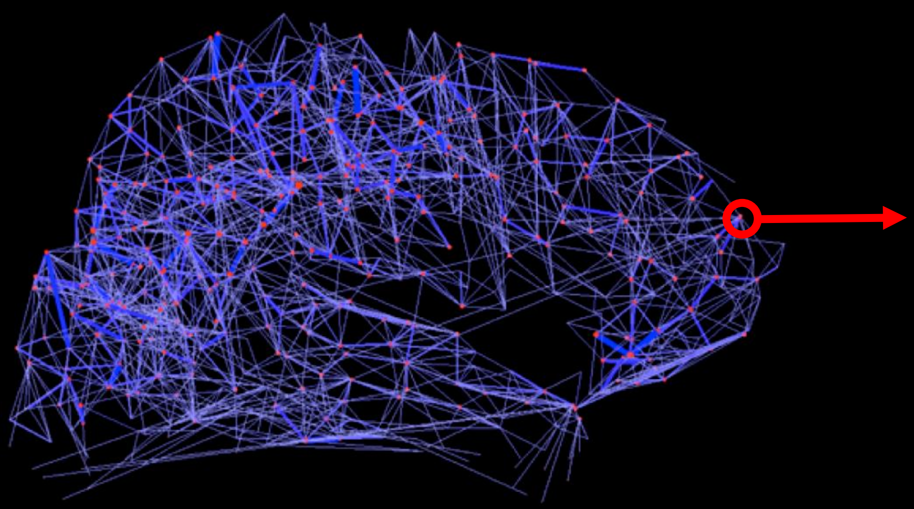
- **Modeling relations from data**
- **Inferring information from data**
- **Classifying data**
- **Predicting outcomes from data**

- **Identifying computational principles from the brain**
- **Understanding neuronal coding and behaviour**

Other definition (emphasizes more the informatics part):

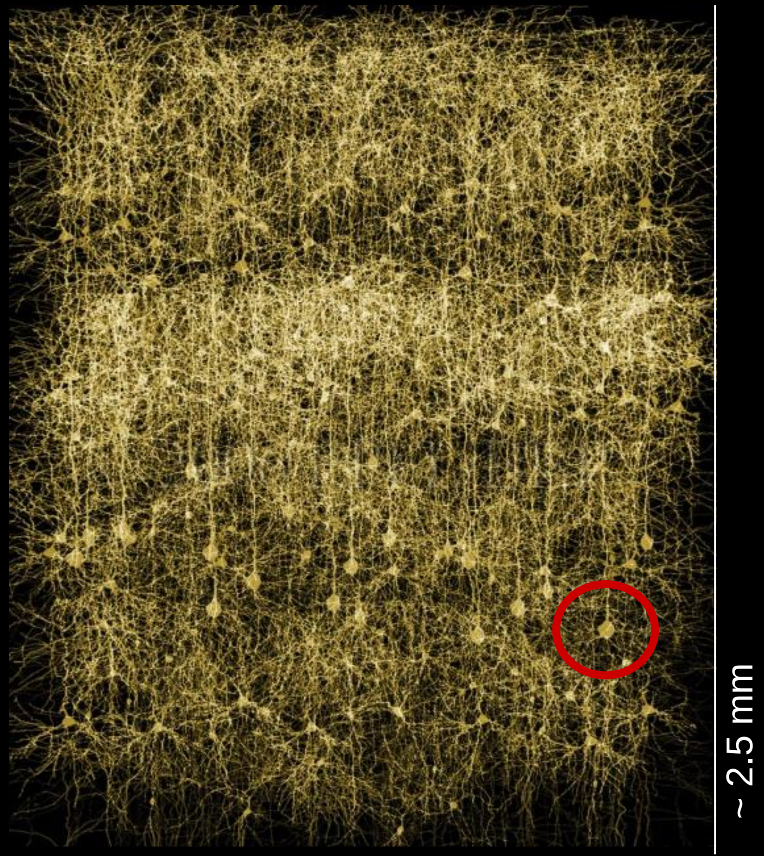
Development of tools for simulating neuronal activity, storing data. It is more like bioinformatics for neuroscience

Macroscopic network



Hagmann, et al. (2008), 'Mapping the structural core of human cerebral cortex' PLoS Biol 6(7): e159

Mesoscopic network

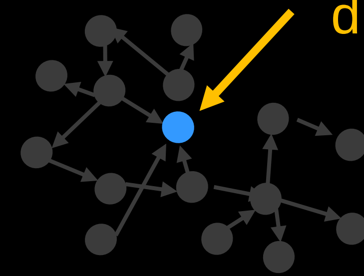


Modified figure from Blue Brain Project

Multiscale complex recurrent network of very heterogeneous elements that is self-organized and delay-coupled

Nested multi-scale encoding model

To what degree is the neuron's response driven by the input or stimulus, and emergent network activity ?



stimulus / drive

stimulus [s(x,y,t)]

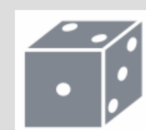
receptive field (i.e. functional)

stimulus drive

d(t)

neuron model (i.e. functional)

p(d(t)) +



output $y(t)$: (point process)

history dependence

Embedding population activity

P(t)

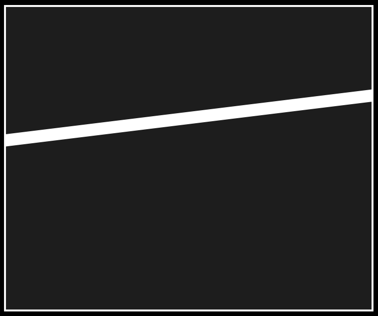
M(t)

Neuronal mass activity

- R. Haslinger, G. Pipa, B. Lima, W. Singer, E. N. Brown and S. Neuenschwander, 'Context Matters: The Illusive Simplicity of Macaque V1 Receptive Fields ' (accepted in PLOS ONE)
- S. V. David, W. E. Vinje, J. L. Gallant, 'Natural Stimulus Statistics Alter the Receptive Field Structure of V1 Neurons', *The Journal of Neuroscience*, August 4, 2004 • 24(31):6991–7006 • 6991

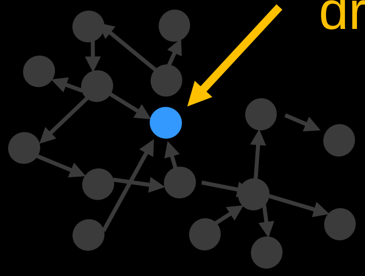
Classical CRF

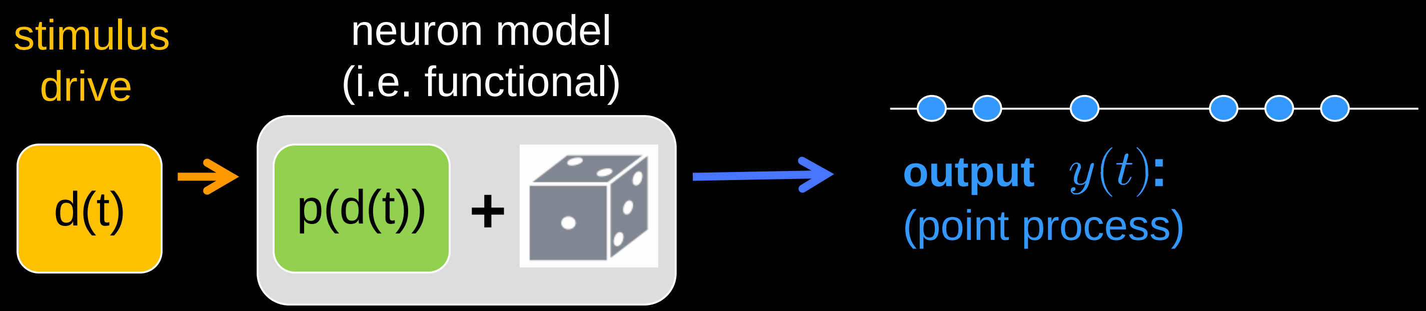
moving bars



To what degree is the neuron's response driven by the input or stimulus, and emergent network activity ?

stimulus / drive

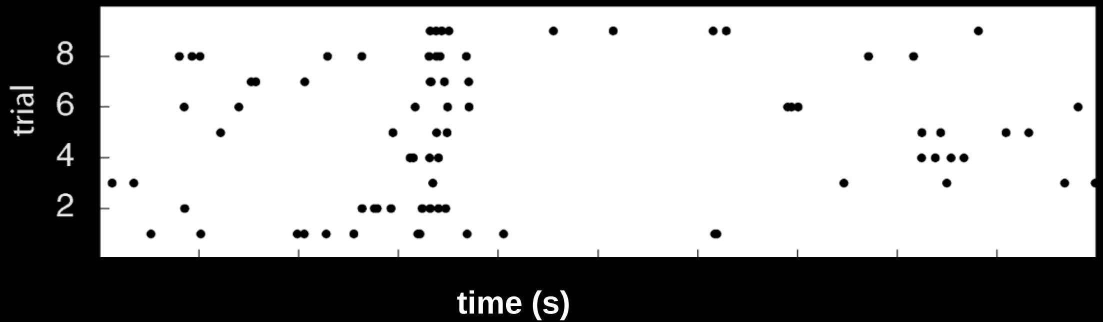




- G. Pipa, S. Neuenschwander, B. Lima, Z. Chen, E.N. Brown, 'Tomographic Reconstruction of Receptive fields', accepted in *Neural Computation*
- R. Haslinger, G. Pipa, B. Lima, W. Singer, E. N. Brown and S. Neuenschwander, 'Context Matters: The Illusive Simplicity of Macaque V1 Receptive Fields ' (accepted in PLOS ONE)

1. Point Process GLM: inhomogeneous Poisson model

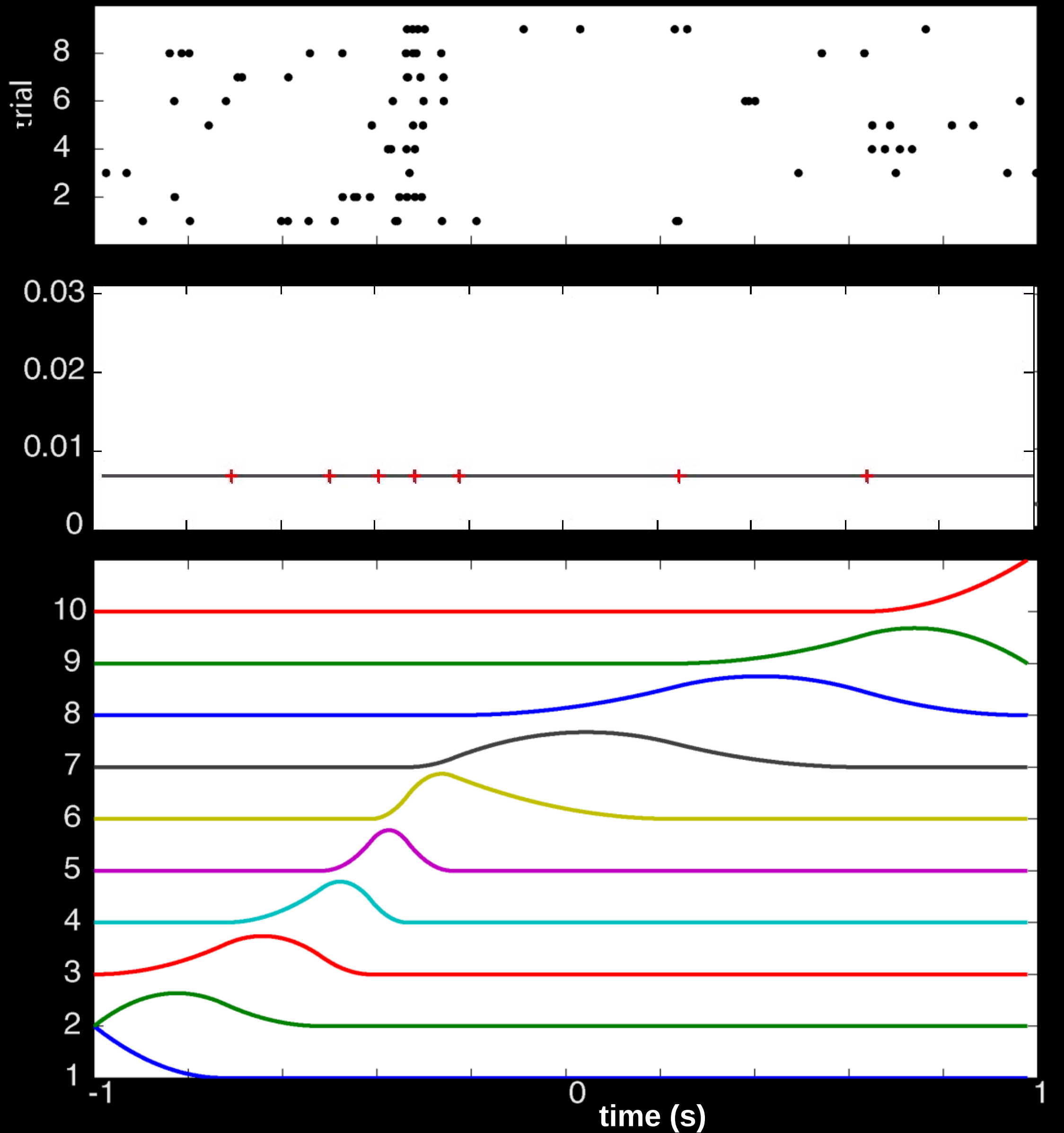
$$\lambda(t_k|f_\beta)_{GLM} = e^{\psi_{\text{stim}}(t_k)}$$



Modeling of the encoding

1. Point Process GLM: inhomogeneous Poisson model

$$\lambda(t_k|f_\beta)_{GLM} = e^{\psi_{\text{stim}}(t_k)}$$



Model is based on cubic splines

Dataset:

1 cell and 9 trials for 1 moving directions β

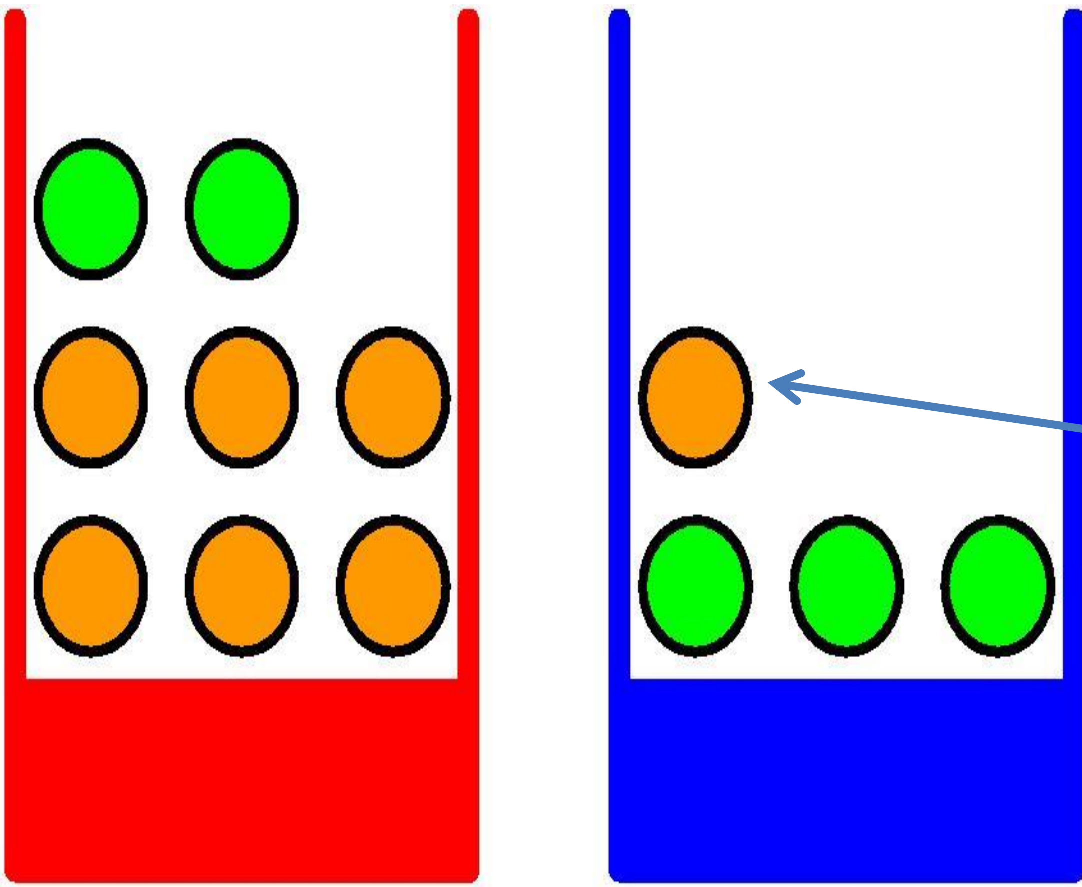
Neuroinformatics Lecture

Topics:

Probabilities and Bayes’

Apples and Oranges

2 x 2 Classes (color x fruit)

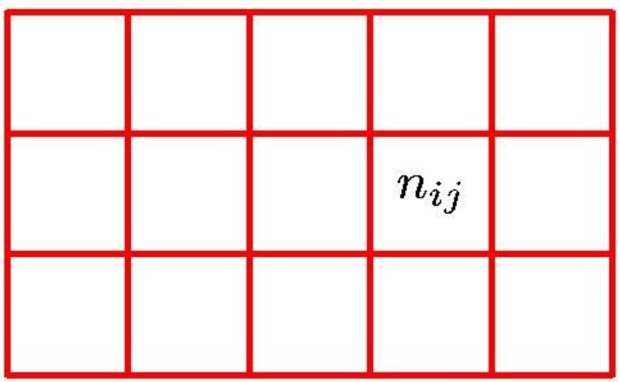


	Blue	Red
Appels	3	2
Oranges	1	6

frequency of occurrence (here blue and the fruit orange)

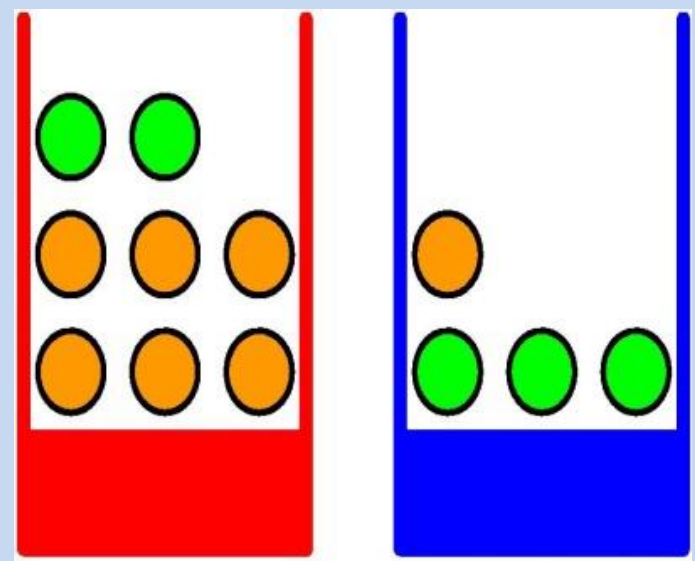
Classes of events x and y i.e. :

Classes x: Color: Red or Blue
Classes y: Fruit: Apple or Orange



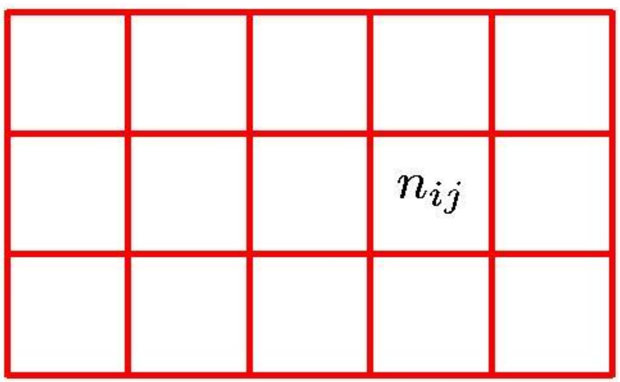
n_{ij} are the Counts

	Blue	Red
Appels	3	2
Oranges	1	6



Classes of events x and y i.e. :

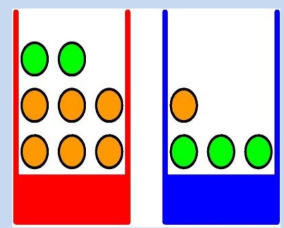
Classes x: Color: Red or Blue
Classes y: Fruit: Apple or Orange



Marginal Probability

(Probability of X=x_i irrespective of class y)

$$p(X = x_i)$$



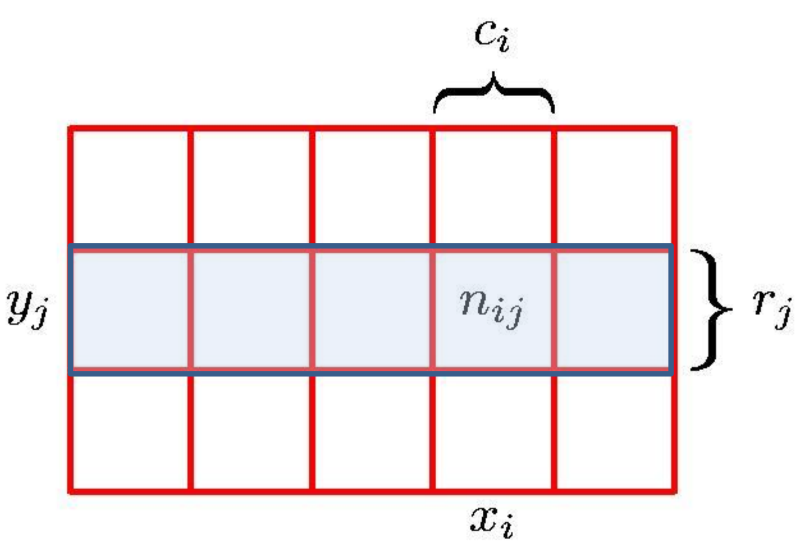
	Blue	Red	
Appels	3	2	5
Oranges	1	6	7
	4	8	N=12

What does our intuition say ?

- Probability of X=Red irrespective of the fruit ?
- Probability of X=blue irrespective of the fruit ?
- Probability of Y= apple irrespective of the color ?
- Probability of Y= orange irrespective of the color ?

Classes of events x and y i.e. :

Classes x: Color: Red or Blue
Classes y: Fruit: Apple or Orange

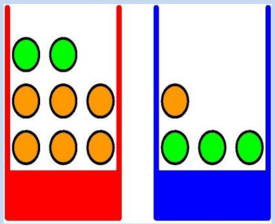


$$r_j = \sum_{i}^{I} n_{i,j}$$

Marginal Probability

(Probability of $X=x_i$ irrespective of class y)

$$p(X = x_i)$$



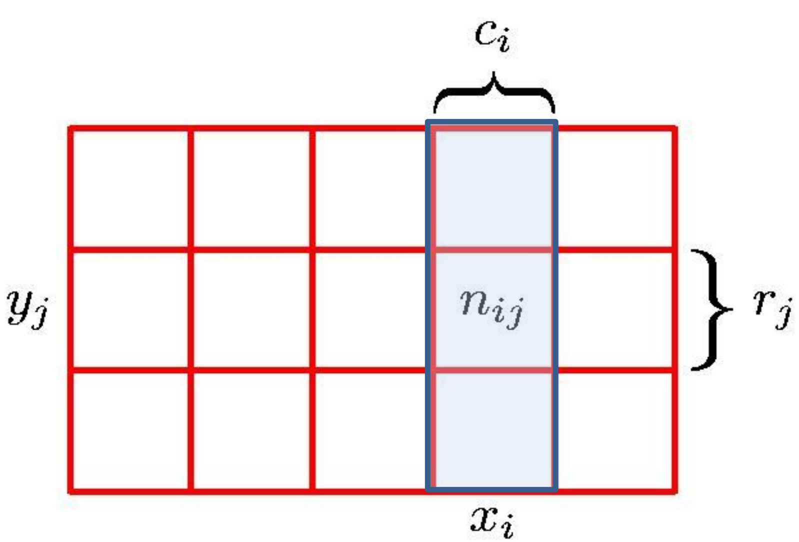
	Blue	Red	
Appels	3	2	5
Oranges	1	6	7
	4	8	N=12

What does our intuition say ?

- Probability of X=Red irrespective of the fruit ?
- Probability of X=blue irrespective of the fruit ?
- Probability of Y= apple irrespective of the color ?
- Probability of Y= orange irrespective of the color ?

Classes of events x and y i.e. :

Classes x: Color: Red or Blue
Classes y: Fruit: Apple or Orange

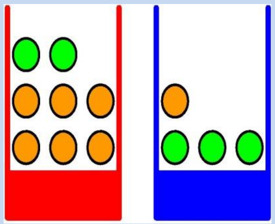


$$c_i = \sum_j n_{i,j}$$

Marginal Probability

(Probability of $X=x_i$ irrespective of class y)

$$p(X = x_i)$$



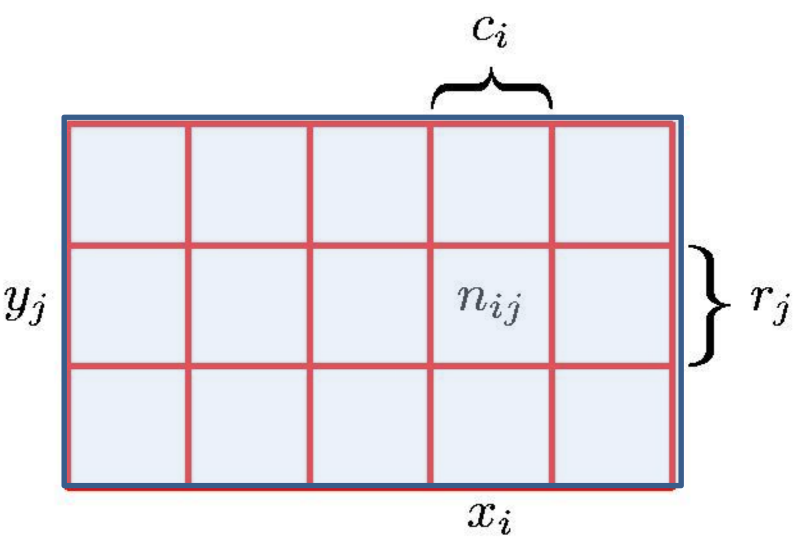
	Blue	Red	
Appels	3	2	5
Oranges	1	6	7
	4	8	N=12

What does our intuition say ?

- Probability of X=Red irrespective of the fruit ?
- Probability of X=blue irrespective of the fruit ?
- Probability of Y= apple irrespective of the color ?
- Probability of Y= orange irrespective of the color ?

Classes of events *x* and *y* i.e. :

Classes x: Color: Red or Blue
Classes y: Fruit: Apple or Orange

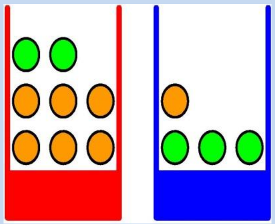


$$N = \sum_{i,j} n_{i,j}$$

Marginal Probability

(Probability of $X=x_i$ irrespective of class y)

$$p(X = x_i)$$



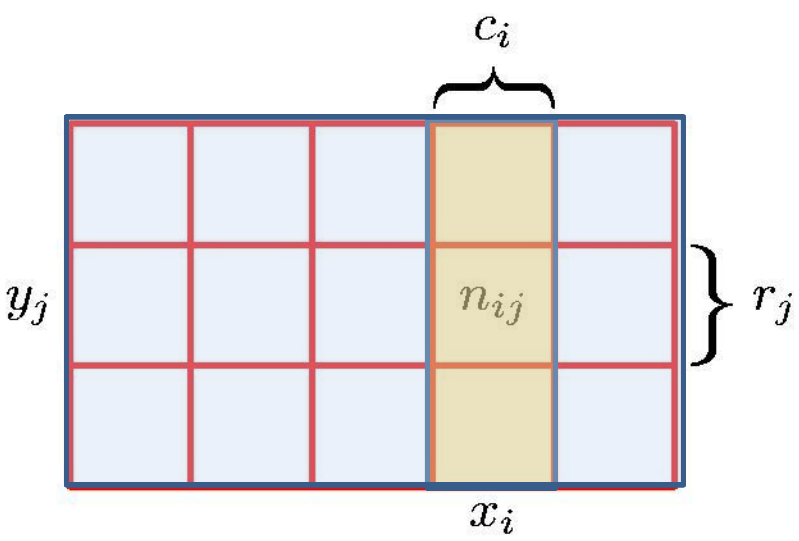
	Blue	Red	
Appels	3	2	5
Oranges	1	6	7
	4	8	N=12

What does our intuition say ?

- Probability of X=Red irrespective of the fruit ?
- Probability of X=blue irrespective of the fruit ?
- Probability of Y= apple irrespective of the color ?
- Probability of Y= orange irrespective of the color ?

Classes of events x and y i.e. :

Classes x: Color: Red or Blue
Classes y: Fruit: Apple or Orange



$$N = \sum_{i,j} n_{i,j} \qquad c_i = \sum_j n_{i,j}$$

Marginal Probability

(Probability of $X=x_i$ irrespective of class y)

$$p(X = x_i)$$

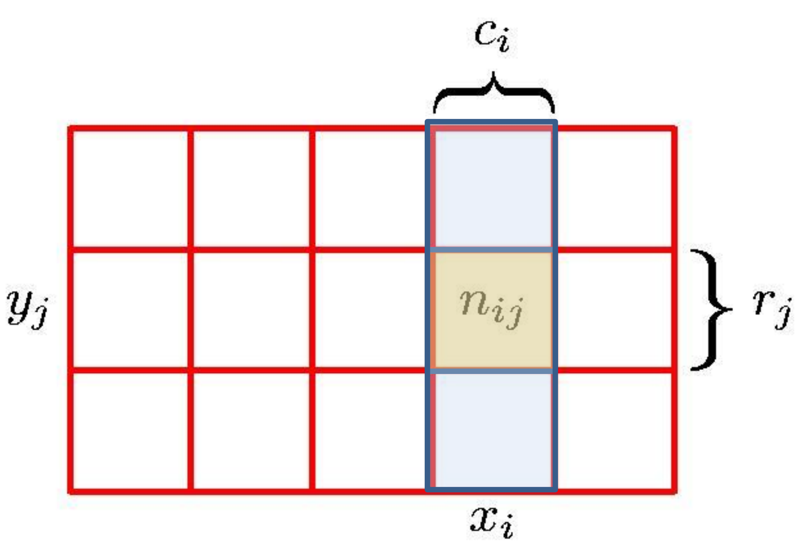
X is a random variable.

$$p(X = x_i) =$$

That is the probability that the random variable X takes the value x_i

Classes of events x and y i.e. :

Classes x: Color: Red or Blue
Classes y: Fruit: Apple or Orange

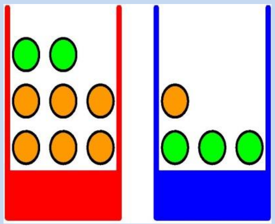


$$N = \sum_{i,j} n_{i,j} \qquad c_i = \sum_j n_{i,j}$$

Conditional Probability

(Probability of $Y=y_j$ given class $X=x_i$)

$$p(Y = y_j | X = x_i)$$



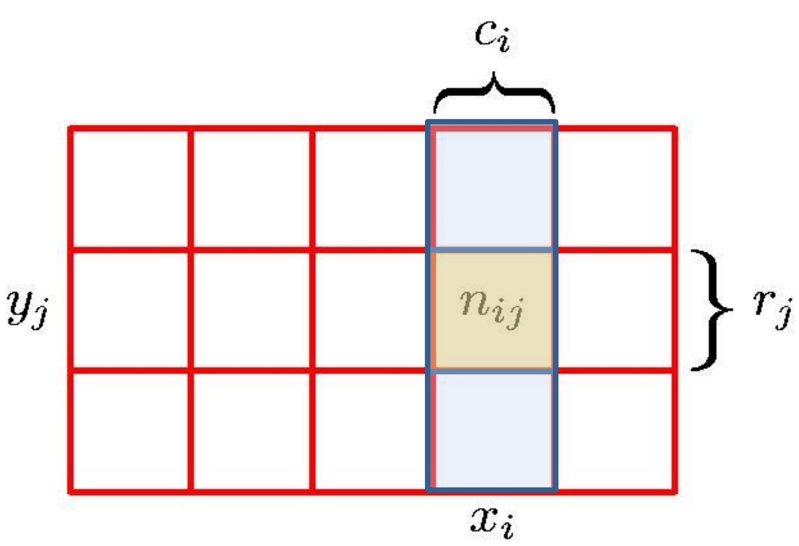
	Blue	Red	
Appels	3	2	5
Oranges	1	6	7
	4	8	N=12

What does our intuition say ?

- Probability of X=Red given fruit=apple ?
- Probability of X=Red given fruit=orange?
- Probability fruit=apple given X=Blue?
- Probability fruit=organge given X=Blue?

Classes of events x and y i.e. :

Classes x: Color: Red or Blue
Classes y: Fruit: Apple or Orange



$$N = \sum_{i,j} n_{i,j} \qquad c_i = \sum_j n_{i,j}$$

Marginal Probability

(Probability of $X=x_i$ irrespective of class y)

$$p(X = x_i) = \frac{c_i}{N}.$$

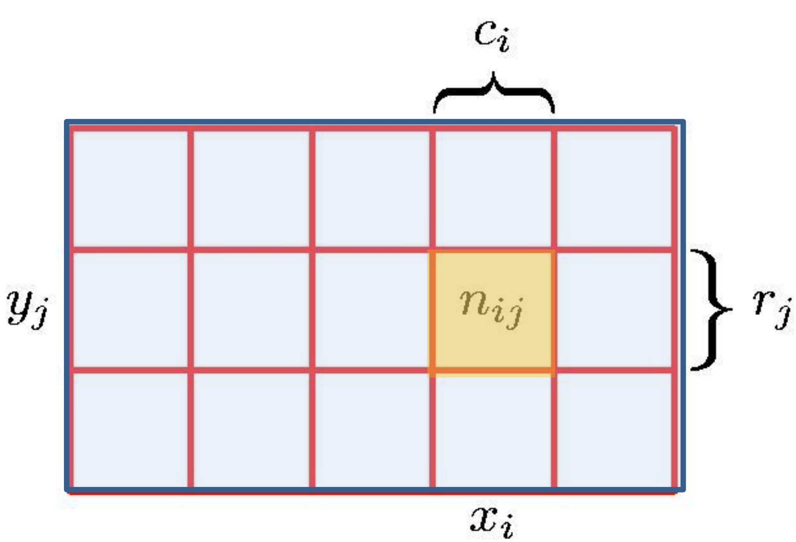
Conditional Probability

(Probability of $Y=y_j$ given class $X=x_i$)

$$p(Y = y_j | X = x_i) = \frac{n_{ij}}{c_i}$$

Classes of events x and y i.e. :

Classes x: Color: Red or Blue
Classes y: Fruit: Apple or Orange

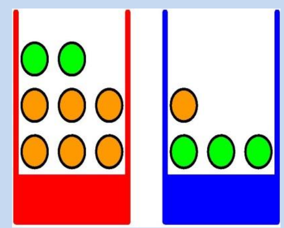


$$N = \sum_{i,j} n_{i,j} \qquad c_i = \sum_j n_{i,j}$$

Joint Probability

(Probability of X=x and Y=y)

$$p(X = x_i, Y = y_j)$$



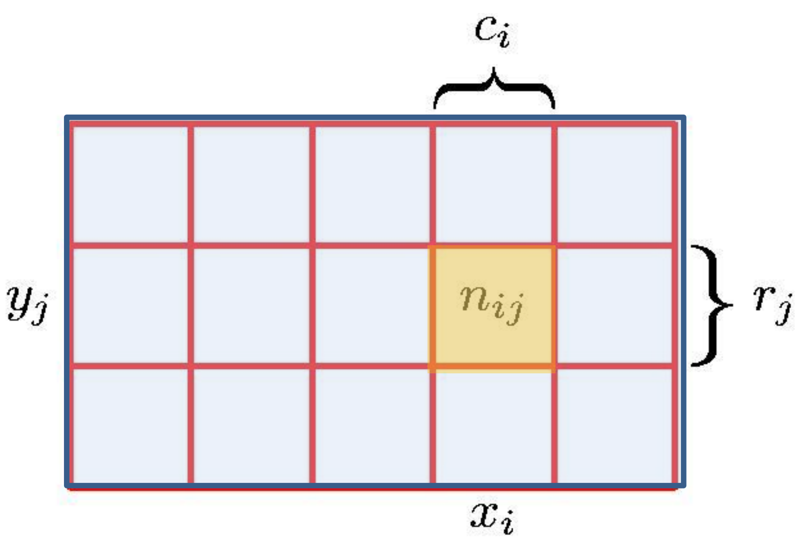
	Blue	Red	
Appels	3	2	5
Oranges	1	6	7
	4	8	N=12

What does our intuition say ?

- Probability of X=Red and fruit=apple ?
- Probability of X=Red and fruit=orange?
- Probability of X=Blue and fruit=apple ?
- Probability of X=Blue and fruit=orange?

Classes of events x and y i.e. :

Classes x: Color: Red or Blue
Classes y: Fruit: Apple or Orange



$$N = \sum_{i,j} n_{i,j} \qquad c_i = \sum_j n_{i,j}$$

Marginal Probability

(Probability of X=x irrespective of class y)

$$p(X = x_i) = \frac{c_i}{N}$$

Conditional Probability

(Probability of Y=y given class X=x)

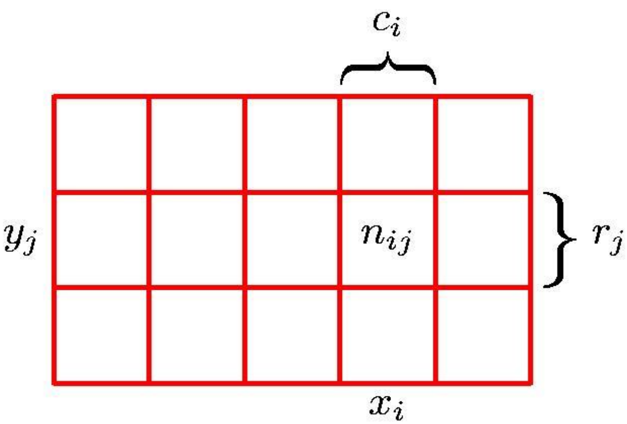
$$p(Y = y_j | X = x_i) = \frac{n_{ij}}{c_i}$$

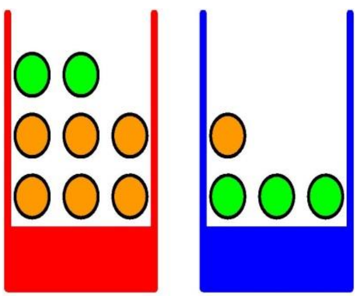
Joint Probability

(Probability of X=x and Y=y)

$$p(X = x_i, Y = y_j) =$$

Probability Theory





$$N = \sum_{i,j} n_{i,j}$$

$$c_i = \sum_{j} n_{i,j}$$

	Blue	Red	
App.	3	2	5
Oran.	1	6	7
	4	8	N=12

Marginal Probability

(Probability of X=x irrespective of class y)

$$p(X = x_i) = \frac{c_i}{N}.$$

Conditional Probability

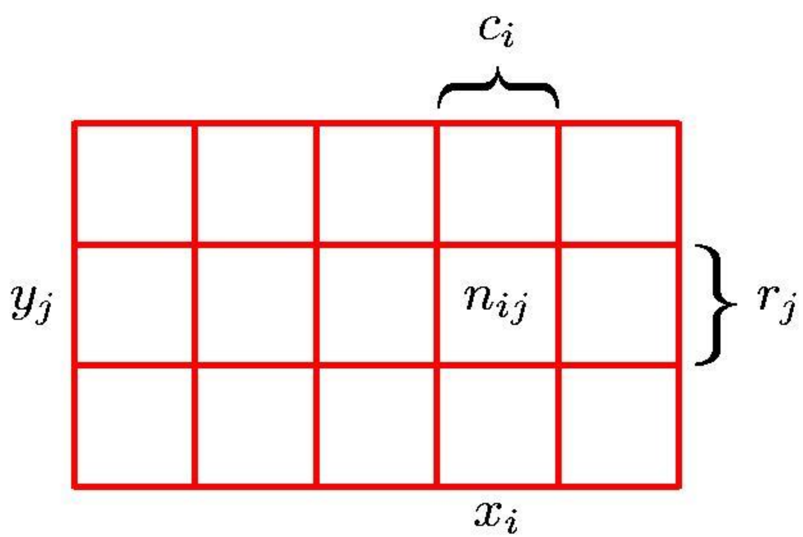
(Probability of Y=y given class X=x)

$$p(Y = y_j | X = x_i) = \frac{n_{ij}}{c_i}$$

Joint Probability

(Probability of X=x and Y=y)

$$p(X = x_i, Y = y_j) = \frac{n_{ij}}{N}$$



marginal probability

Sum Rule

$$p(X = x_i) = \frac{c_i}{N} =$$

joint probability

Product Rule

$$p(X = x_i, Y = y_j) = \frac{n_{ij}}{N}$$

joint probability

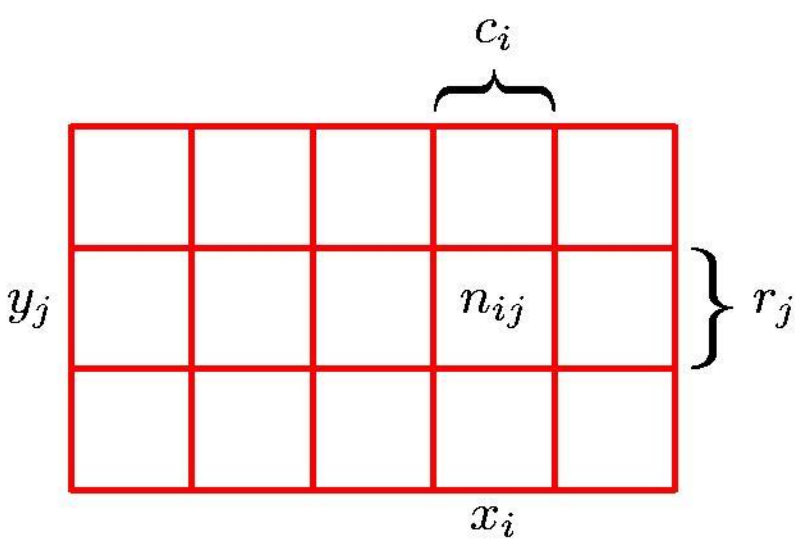
conditional probability

marginal probability

Sum Rule $$p(X) = \sum_Y p(X, Y)$$

Product Rule $$p(X, Y) = p(Y|X)p(X)$$

Questionnaire



Instructions: Solve the following problems regarding the requested order during your lecture.

1. Use c_i, r_j, and $n_{i,j}$ to compute the following probabilities using their relative frequencies :

1. joint probability $p(X = x_i, Y = y_i)$ using $n_{i,j}$:
2. marginal probability $p(X = x_i)$ using c_i:
3. marginal probability $p(X = x_i)$ using $n_{i,j}$:

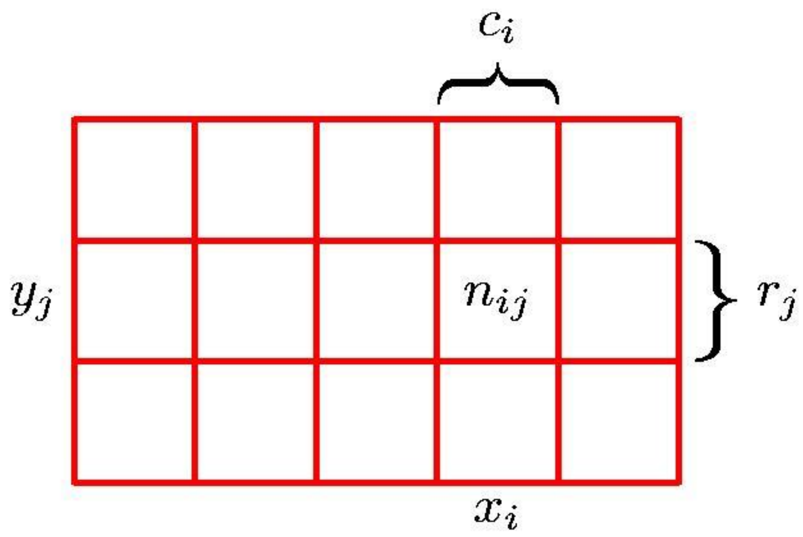
Start

Timer

Stop

freemusicarchive.org/music/Kai_Engel/

Questionnaire



marginal probability

Sum Rule

$$p(X = x_i) = \frac{c_i}{N} = \frac{1}{N}\sum_{j=1}^{L} n_{ij}$$

$$= \sum_{j=1}^{L} p(X = x_i, Y = y_j)$$

joint probability

Product Rule

$$p(X = x_i, Y = y_j) = \frac{n_{ij}}{N} = \frac{n_{ij}}{c_i} \cdot \frac{c_i}{N}$$

$$= p(Y = y_j | X = x_i) p(X = x_i)$$

joint probability

conditional probability

marginal probability