

**10-601**

# Machine Learning

<http://www.cs.cmu.edu/afs/cs/academic/class/10601-f10/index.html>

# Course data

- All up-to-date info is on the course web page:

<http://www.cs.cmu.edu/afs/cs/academic/class/10601-f10/index.html>

- Instructor:

- Ziv Bar-Joseph

- TAs:

- Anthony Gitter

- Hai-Son Le

- Yang Xu

- See web page for contact info, office hours, etc.

- Mailing list - send email to: [yx1@cs.cmu.edu](mailto:yx1@cs.cmu.edu)

- Follow us on Twitter: ML10601

- 8/23/10 – Intro to ML and probability
- 8/25/10 – Density estimation, classification theory
- 8/30/10 – Classification
- 09/01/10 – Bayes and Naïve Bayes classifiers / PS1
- 9/8 (Wednesday): No class (Matlab recitation)
- 09/15/10 – SVM 1 / PS1 in PS2 out
- 09/20/10 – SVM 2
- 09/22/10 – Boosting
- 09/27/10 – Learning theory 1
- 09/29/10 – Learning theory 2 / PS2 in PS3 out
- 10/04/10 – Decision trees
- 10/06/10 – NN
- 10/11/10 - Hierarchical clustering
- 10/13/10 – Kmeans and Gaussian mixtures / PS3 in PS4 out
- 10/20/10 – Model selection, feature selection
- 10/18/10 – semi supervised learning / Project proposal
- 10/25/10 – ML in industry 1
- 10/27/10 – ML in industry 2 / PS4 in
- 11/3 (Wednesday): Midterm (1:30-3:30)
- 11/10/10 – HMM2 structure learning / Project progress report
- 11/15/10 – MDPs
- 11/17/10 – PCA, SVD / PS5 in
- 11/22/10 – Graph clustering? RL?
- 11/24/10 – Thanksgiving
- 12/1 (Wednesday): Poster session in the afternoon (class as usual)

**Intro and classification  
(A.K.A. ‘supervised learning’)**

**Clustering  
(‘Unsupervised learning’)**

**Probabilistic representation  
and modeling (‘reasoning  
under uncertainty’)**

**Applications  
of ML**

# Grading

- **5 Problem sets - 40%**
- **Project - 30%**
- **Midterm - 25%**
- **Class participation - 5%**

# Class assignments

- 5 Problem sets
  - Each containing both theoretical and programming assignments
- Projects
  - Groups of 1 or 2
  - Implement and apply an algorithm discussed in class to a new domain
  - Extend algorithms discussed in class in various directions
  - New theoretical results (for example, for a new setting of a problem)
  - More information on website
- Recitations
  - Monday, 5-6:20pm, NSH 1305
  - Expand on material learned in class, go over problems from previous classes etc.

# What is Machine Learning?

Easy part: Machine

Hard part: Learning

- Short answer: Methods that can help generalize information in observed data so that it can be used to make better decisions in the future

# What is Machine Learning?

Longer answer: The term Machine Learning is used to characterize a number of different approaches for generalizing from observed data:

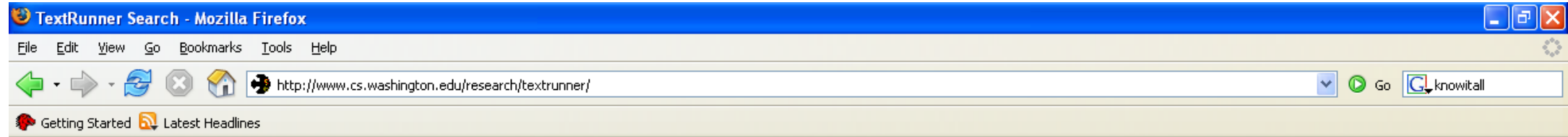
- Supervised learning
  - Given a set of features and labels learn a model that will predict a label to a new feature set
- Unsupervised learning
  - Discover patterns in data
- Reasoning under uncertainty
  - Determine a model of the world either from samples or as you go along
- Active learning
  - Select not only model but also which examples to use

# Paradigms of ML

- Supervised learning
  - Given  $D = \{X_i, Y_i\}$  learn a model (or function)  $F: X_k \rightarrow Y_k$
- Unsupervised learning
  - Given  $D = \{X_i\}$  group the data into  $Y$  classes using a model (or function)  $F: X_i \rightarrow Y_j$
- Reinforcement learning (reasoning under uncertainty)
  - Given  $D = \{\text{environment, actions, rewards}\}$  learn a policy and utility functions:  
  
policy:  $F1: \{e, r\} \rightarrow a$   
utility:  $F2: \{a, e\} \rightarrow R$
- Active learning
  - Given  $D = \{X_i, Y_i\}, \{X_j\}$  learn a function  $F1: \{X_j\} \rightarrow x_k$  to maximize the success of the supervised learning function  $F2: \{X_i, x_k\} \rightarrow Y$



# Web search



## Turing Center KnowItAll Project



## TextRunner Search

TextRunner searches hundreds of millions of assertions extracted from over 100 million Web pages on the topics of nutrition, history of science, and general knowledge, and sorts the results by probability.

Our IJCAI '07 paper on TextRunner is here: [Open Information Extraction from the Web](#)

Example queries:

["What did Thomas Edison invent?"](#)

["What kills bacteria?"](#)

["Johannes Kepler"](#)

Search individual fields:

Argument 1

Predicate

Argument 2

Search

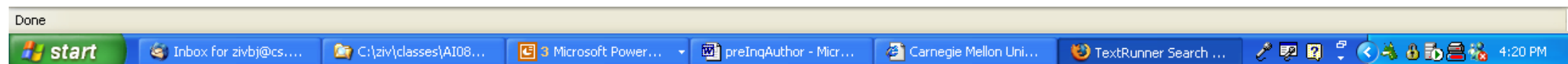
Primarily unsupervised learning

Search query:

Search

[questions/comments/bugs](#)

[Show advanced search options](#)



# Web search, cont'd

**flipdog**™ [find local jobs](#)

Job Title, Keywords: San Francisco

Powered

## Senior Marketing Representative

Crawley Warren Insurance Services, Inc. (San Francisco, California)

---

**Salary:** \$20 to \$30  
**Salary Details:** depending on experience  
**Position Type:** Parttime  
**Ref Code:** 60576596  
**Minimum Education Level:**  
Some College Coursework Completed  
**Minimum Career Level:** Experienced (Non-Manager)

[Save Job to my monster](#) [APPLY NOW](#)

Primarily supervised learning

# Recommender systems

Amazon.com: Recommended for You - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://www.amazon.com/gp/yourstore/ref=pd\_irl\_283155?ie=UTF8&nodeID=283155&rGroup=books&pf\_rd\_p=273115801&pf\_rd\_s=center-2&pf\_rd\_t=101&pf\_rd\_i=2831558

amazon.com Hello, Ziv Bar-Joseph. We have recommendations for you. (Not Ziv?)

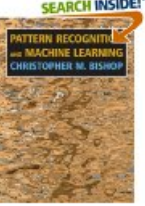
Shop All Departments Search Books

Recommended For You

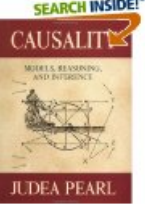
Recommended For You > Books

These recommendations are based on [items you own](#) and more.

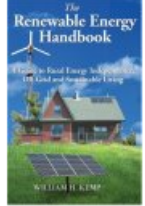
view: All | [New Releases](#) | [Coming Soon](#)

- 


**Pattern Recognition and Machine Learning (Information Science and Statistics)**  
by Christopher M. Bishop (Oct 1, 2007)  
Average Customer Review: ★★★★★ (38)  
In Stock  
List Price: \$84.95  
Price: **\$62.60**  
56 used & new from \$56.64

I own it  Not interested x|★★★★★ Rate it  
Recommended because you purchased **Learning in Graphical Models** and more ([Fix this](#))
- 

**Causality: Models, Reasoning, and Inference**  
by Judea Pearl (Mar 13, 2000)  
Average Customer Review: ★★★★★ (12)  
In Stock  
List Price: \$50.00  
Price: **\$38.50**  
26 used & new from \$32.01

I own it  Not interested x|★★★★★ Rate it  
Recommended because you purchased **Probabilistic Reasoning in Intelligent Systems** and more ([Fix this](#))
- 

**The Renewable Energy Handbook: A Guide to Rural Energy Independence, Off-Grid and Sustainable Living**  
by William H. Kemp (April 1, 2006)  
Average Customer Review: ★★★★★ (16)  
In Stock  
List Price: \$29.95  
Price: **\$19.77**  
40 used & new from \$18.25

I own it  Not interested x|★★★★★ Rate it  
Recommended because you purchased **Wind Power, Revised Edition** and more ([Fix this](#))
- 

**Learning Bayesian Networks (Artificial Intelligence)**  
by Richard E. Neapolitan (April 6, 2003)  
Average Customer Review: ★★★★★ (2)

http://www.amazon.com/Pattern-Recognition-Learning-Information-Statistics/dp/0387310738/ref=pd\_ys\_irl\_b\_1?pf\_rd\_p=258372101&pf\_rd\_s=center-1&pf\_rd\_t=1501&pf\_rd\_i=list&pf\_rd\_m=ATVPDKIKX0DER&pf\_rd\_r=1BQMM558P495ESDQ98HP

start | Inbox for zivbj@cs... | C:\ziv\classes\AI08... | 3 Microsoft Power... | preInqAuthor - Micr... | Carnegie Mellon Uni... | Amazon.com: Reco... | 4:28 PM

Primarily supervised learning

# Grand and Urban Challenges road race

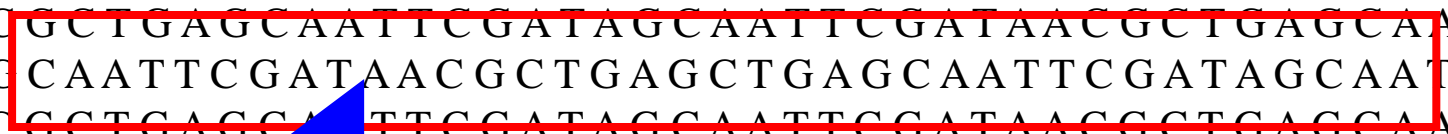
Supervised and  
reinforcement learning

# Helicopter control

Reinforcement learning

# Biology

ACGCTGAGCAATTCGATAGCAATTCGATAACGCTGAGCAACGCTGAGCAATTCGATAGCAATTC  
GATAACGCTGAGCAATTCGGATAACGCTGAGCAATTCGATAGCAATTCGATAACGCTGAGCAACG  
CTGAGCAATTCGATAGCAATTCGATAACGCTGAGCAATTCGGATATCGATAGCAATTCGATAAATC  
GGATAACGCTGAGCAATTCGATAGCAATTCGATAACGCTGAGCAACGCTGAGCAATTCGATAGC  
AATTCGATAACGCTGAGCAATTCGGATATCGATAGCAATTCGATAACGCTGAGCAACGCTGAGCA  
ATTCGATAGCAATTCGATAACGCTGAGCAATTCGGATAACGCTGAGCAATTCGATAGCATTTCGAT  
AACGCTGAGCAACGCTGAGCAATTCGATAGCAATTCGATAACGCTGAGCAATTCGGATAACGCTG  
AGCAATTCGATAGCAATTCGATAACGCTGAGCTGAGCAATTCGATAGCAATTCGATAACGCTGA  
GCAATTCGGATAACGCTGAGCAATTCGATAGCAATTCGATAACGCTGAGCAACGCTGAGCAATTC  
GATAGCAATTCGATAGCAATTCGATAACGCTGAGCAACGCTGAGCAATTCGATAGCAATTCGAT  
AGCAATTCGATAACGCTGAGCAATTCGGATAACGCTGAGCAATTCGATAGCAATTCGATAACGCT  
GAGCAACGCTGAGCAATTCGATAGCAATTCGATAACGCTGAGCAATTCGGATATCGATAGCAATT  
CGATAACGCTGAGCAACGCTGAGCAATTCGATAGCAATTCGATAACGCTGAGCAATTCGGATAAC  
GCTGAGCAATTCGATAGCAATTCGATAACGCTGAGCAATTCGATAGCAATTCGATAACGCTGAG  
CTGATAGCAATTCGATAACGCTGAGCAACGCTGAGCAACGCTGAGCA  
ATTCGGATATCGATAGCAATTCGATAACGCTGAGCA  
ACGCTGAGCAATTCGATAGCAATTCGATAACGCTGAGCAATTCGGATAACGCTGAGCAATTCGAT  
AGCATTTCGATAACGCTGAGCAACGCTGAGCAATTCGATAGCAATTCGATCGGATAACGCTGAGC  
AATTCGATAGCAATTCGATAACGCTGAGCTGAGCAATTCGATAGCAATTCGATAACGCTGAGCA  
ATCGGATAACGCTGAGCAATTCGATAGCA  
AGCAATTCGATAACGCTGAGCAATTCGGAT  
GCAATTCGATAGCAATTCGATAACGCTGA  
GATAACGCTGAGCAACGCTGAGCAATTCG  
CTGAGCAATTCGATAGCAATTCGATAACG  
TGAGCAATTCGGATAACGCTGAGCAATTCGATAGCAATTCGATAACGCTGAGCAACGCTGAGCAA  
TTCGATAGCAATTCGATAGCAATTCGATAGCAATTCGATAACGCTGAGCAACGCTGAGCAATTC  
GATAGCAATTCGATAACGCTGAGCAATTCGGATAACGCTGAGCAATTCGATAGCAATTCGATAAC  
GCTGAGCAACGCTGAGCAATTCGATAGCAATTCGATAACGCTGAGCAATTCGGATATCGATAGCA  
ATTCGATAACGCTGAGCAACGCTGAGCAATTCGATAGCAATTCGATAACGCTGAGCAATTCGGAT  
AACGCTGAGCAATTCGATAGCAATTCGATAACGCTGAGCTGAGCAATTCGATAGCAATTCGATA  
ACGCTGAGCAATTCGGA



**Which part is the gene?**

Supervised and  
unsupervised learning (can  
also use active learning)

THE WALL STREET JOURNAL. Digital Network WSJ.com MarketWatch BARRON'S All Things Digital. FINS SmartMoney More

Tuesday, July 13, 2010 New York 95° | 75°

THE WALL STREET JOURNAL. | **TECH**

U.S. Edition Today's Paper Video Blogs Journal Community Log In

Home World U.S. New York Business Markets **Tech** Personal Finance Life & Style Opinion Careers Real Estate Small Business

Digits Personal Technology All Things Digital What They Know

TOP STORIES IN Technology

1 of 10 Google Agonizes Over Privacy

2 of 10 Google, Verizon Propose Web Rules

3 of 10 H-P Investors Head for Door

Electronic

BUSINESS TECHNOLOGY | JULY 13, 2010

## Letting the Machines Decide

*New Wave of Investment Firms Look to 'Artificial Intelligence' in Trade Decisions*

Article

Video

Comments (81)

MORE IN TECH »

Email Print

Save This

Like 33

+ More

Text

By SCOTT PATTERSON

Wall Street is notorious for not learning from its mistakes. Maybe machines can do better.

That is the hope of an increasing number of investors who are turning to the science of artificial intelligence to make investment decisions.



With artificial intelligence, programmers don't just set up computers to make decisions in response to certain inputs. They attempt to



Supervised and learning, regression analysis

*Most investors trying the approach are using "machine learning," a branch of artificial intelligence in which a computer program analyzes huge chunks of data and makes predictions about the future.*



# Common Themes

- Mathematical framework
  - Well defined concepts based on explicit assumptions
- Representation
  - How do we encode text? Images?
- Model selection
  - Which model should we use? How complex should it be?
- Use of prior knowledge
  - How do we encode our beliefs? How much can we assume?



(brief) intro to probability

# Basic notations

- Random variable
  - referring to an element / event whose status is unknown:  
A = “it will rain tomorrow”
- Domain (usually denoted by  $\Omega$ )
  - The set of values a random variable can take:
    - “A = The stock market will go up this year”: Binary
    - “A = Number of Steelers wins in 2007”: Discrete
    - “A = % change in Google stock in 2007”: Continuous

# Axioms of probability (Kolmogorov's axioms)

A variety of useful facts can be derived from just three axioms:

1.  $0 \leq P(A) \leq 1$
2.  $P(\text{true}) = 1, P(\text{false}) = 0$
3.  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

There have been several other attempts to provide a foundation for probability theory. Kolmogorov's axioms are the most widely used.

# Using the axioms

- How can we use the axioms to prove that:

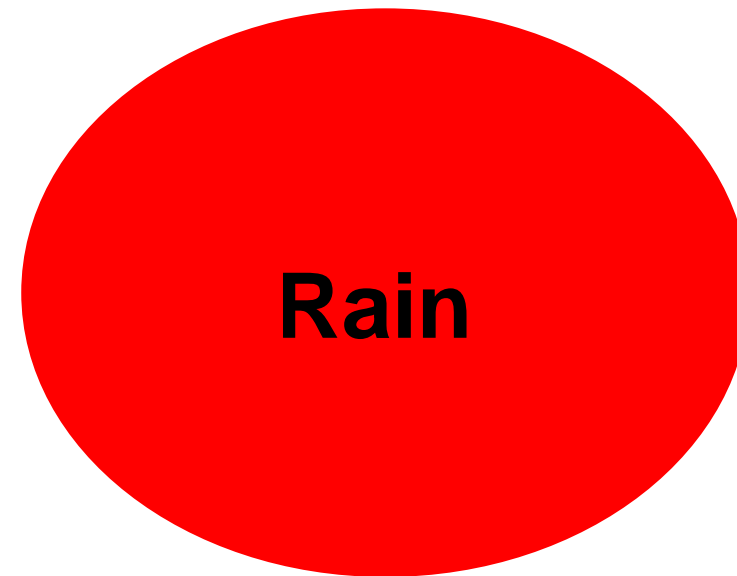
$$P(\neg A) = 1 - P(A)$$

?

# Priors

Degree of belief  
in an event in the  
absence of any  
other information

**No rain**



**Rain**

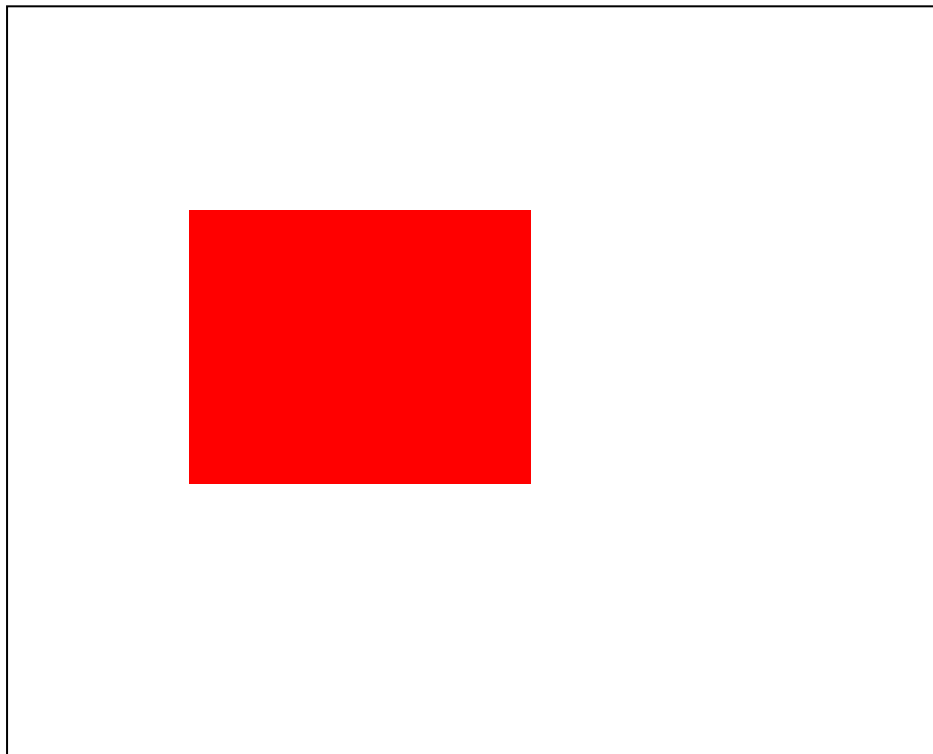
$$P(\text{rain tomorrow}) = 0.2$$

$$P(\text{no rain tomorrow}) = 0.8$$

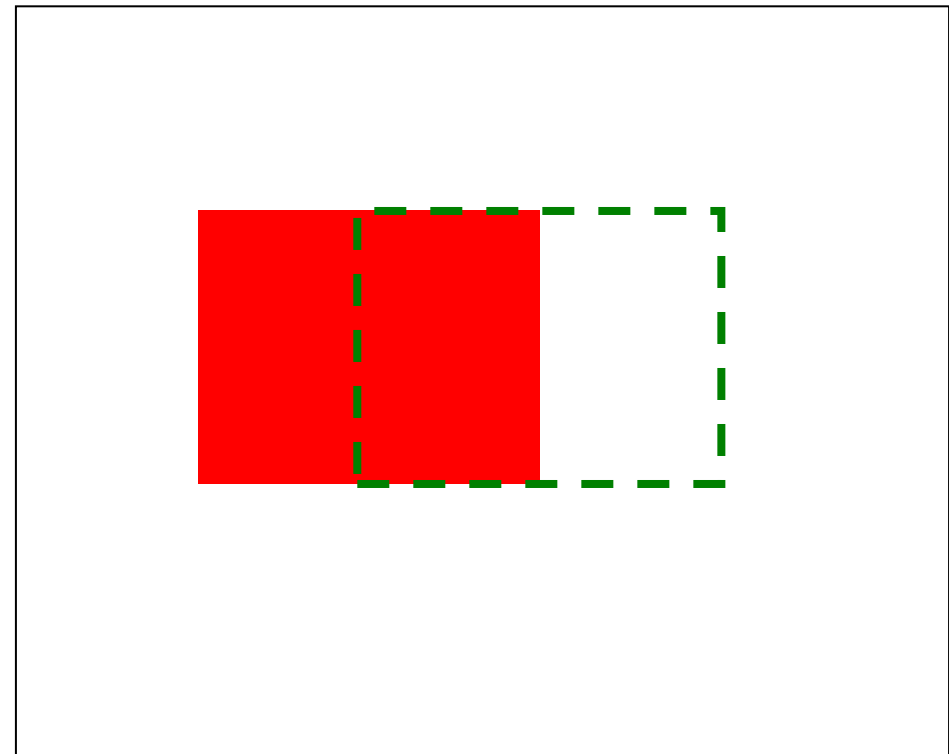
# Conditional probability

- $P(A = 1 \mid B = 1)$ : The fraction of cases where A is true if B is true

$$P(A = 0.2)$$



$$P(A|B = 0.5)$$



# Conditional probability

- In some cases, given knowledge of one or more random variables we can improve upon our prior belief of another random variable
- For example:

$$p(\text{slept in movie}) = 0.5$$

$$p(\text{slept in movie} \mid \text{liked movie}) = 1/4$$

$$p(\text{didn't sleep in movie} \mid \text{liked movie}) = 3/4$$

Slept	Liked
1	0
0	1
1	1
1	0
0	0
1	0
0	1
0	1

# Joint distributions

- The probability that a set of random variables will take a specific value is their joint distribution.
- Notation:  $P(A \wedge B)$  or  $P(A,B)$
- Example:  $P(\text{liked movie, slept})$

If we assume independence then

$$P(A,B)=P(A)P(B)$$

However, in many cases such an assumption maybe too strong (more later in the class)



# Joint distribution (cont)

$P(\text{class size} > 20) = 0.6$

$P(\text{summer}) = 0.4$

$P(\text{class size} > 20, \text{summer}) = ?$

## Evaluation of classes

Size	Time	Eval
30	R	2
70	R	1
12	S	2
8	S	3
56	R	1
24	S	2
10	S	3
23	R	3
9	R	2
45	R	1

# Joint distribution (cont)

$P(\text{class size} > 20) = 0.6$

$P(\text{summer}) = 0.4$

$P(\text{class size} > 20, \text{summer}) = 0.1$

## Evaluation of classes

Size	Time	Eval
30	R	2
70	R	1
12	S	2
8	S	3
56	R	1
24	S	2
10	S	3
23	R	3
9	R	2
45	R	1

# Joint distribution (cont)

$P(\text{class size} > 20) = 0.6$

$P(\text{eval} = 1) = 0.3$

$P(\text{class size} > 20, \text{eval} = 1) = 0.3$

Size	Time	Eval
30	R	2
70	R	1
12	S	2
8	S	3
56	R	1
24	S	2
10	S	3
23	R	3
9	R	2
45	R	1

# Joint distribution (cont)

$P(\text{class size} > 20) = 0.6$

$P(\text{eval} = 1) = 0.3$

$P(\text{class size} > 20, \text{eval} = 1) = 0.3$

## Evaluation of classes

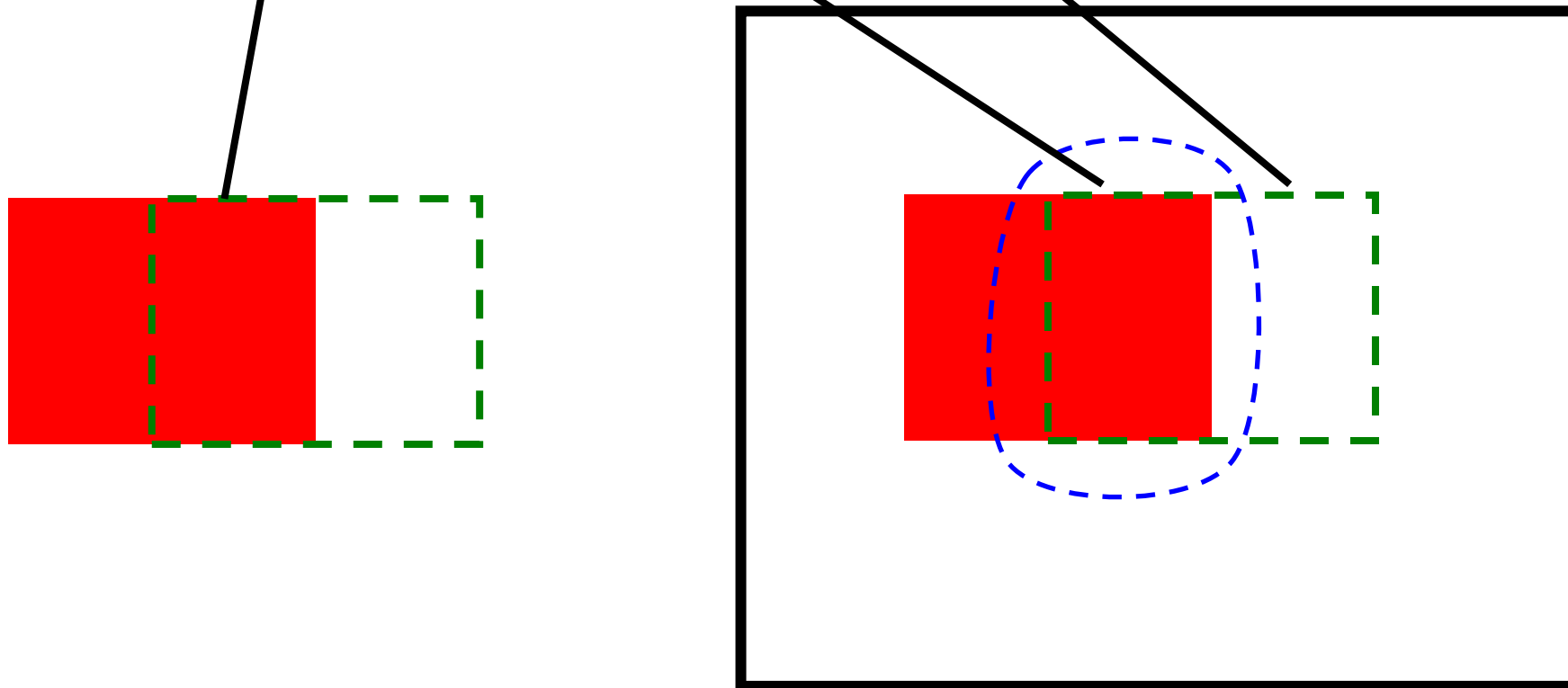
Size	Time	Eval
30	R	2
70	R	1
12	S	2
8	S	3
56	R	1
24	S	2
10	S	3
23	R	3
9	R	2
45	R	1

# Chain rule

- The joint distribution can be specified in terms of conditional probability:

$$P(A,B) = P(A|B)*P(B)$$

- Together with Bayes rule (which is actually derived from it) this is one of the most powerful rules in probabilistic reasoning



# Bayes rule

- One of the most important rules for AI usage.
- Derived from the chain rule:

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

- Thus,

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

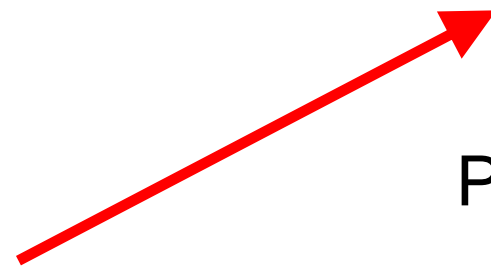


**Thomas Bayes** was an English clergyman who set out his theory of probability in 1764.

# Bayes rule (cont)

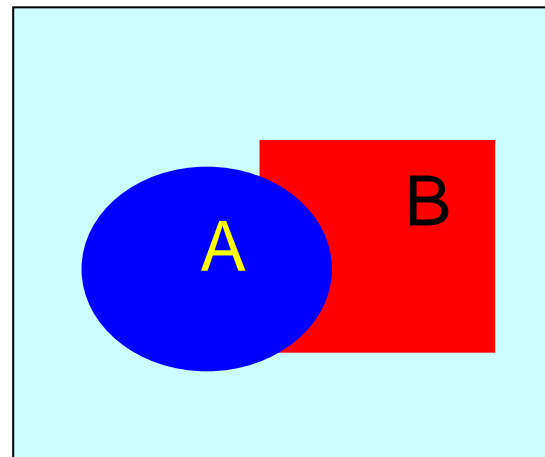
Often it would be useful to derive the rule a bit further:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{P(B|A)P(A)}{\sum_A P(B|A)P(A)}$$

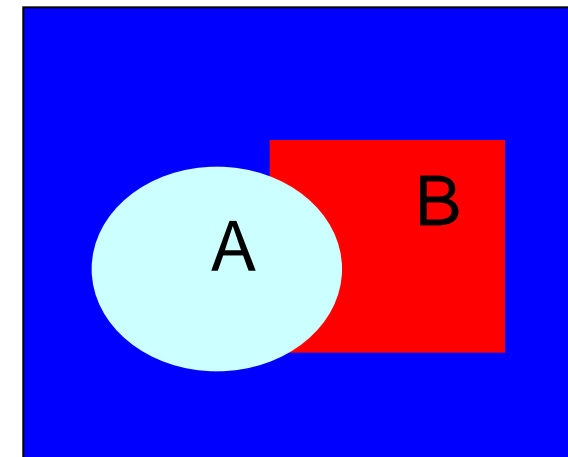


This results from:  
 $P(B) = \sum_A P(B,A)$

$P(B,A=1)$

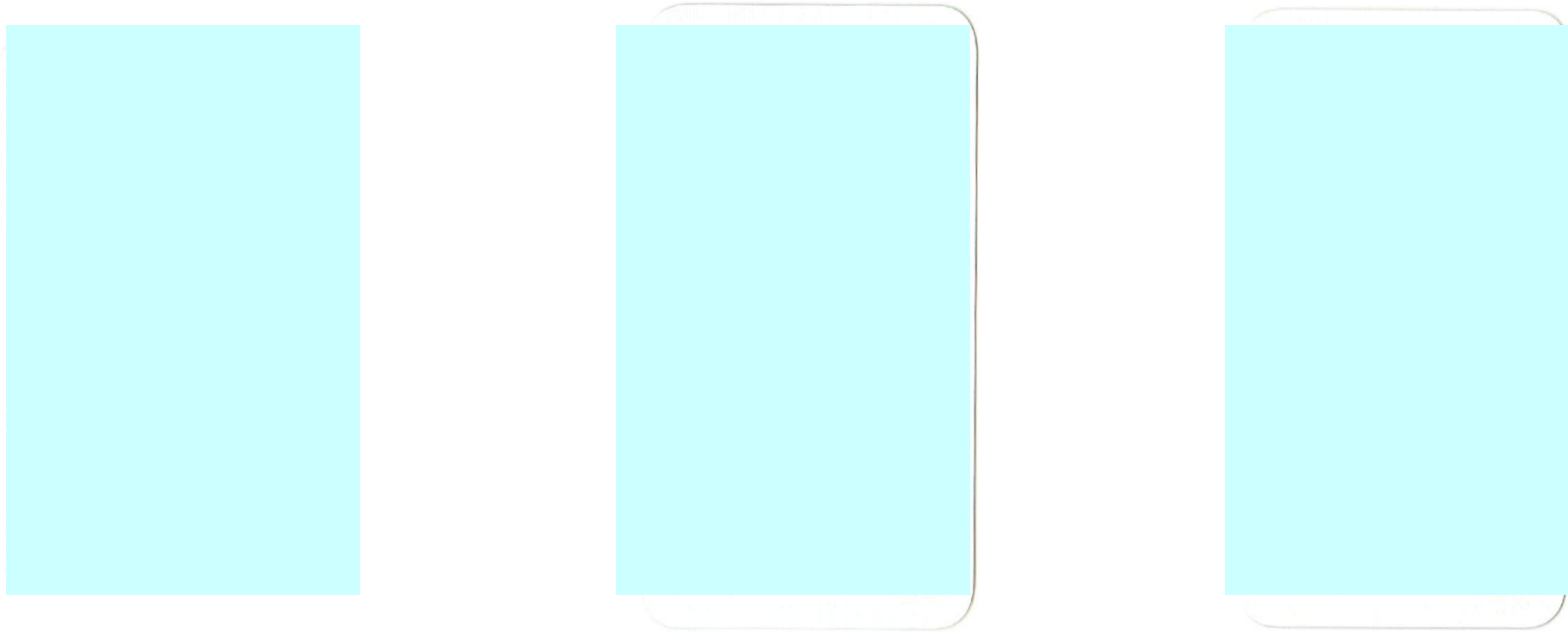


$P(B,A=0)$



# Using Bayes rule

- Cards game:

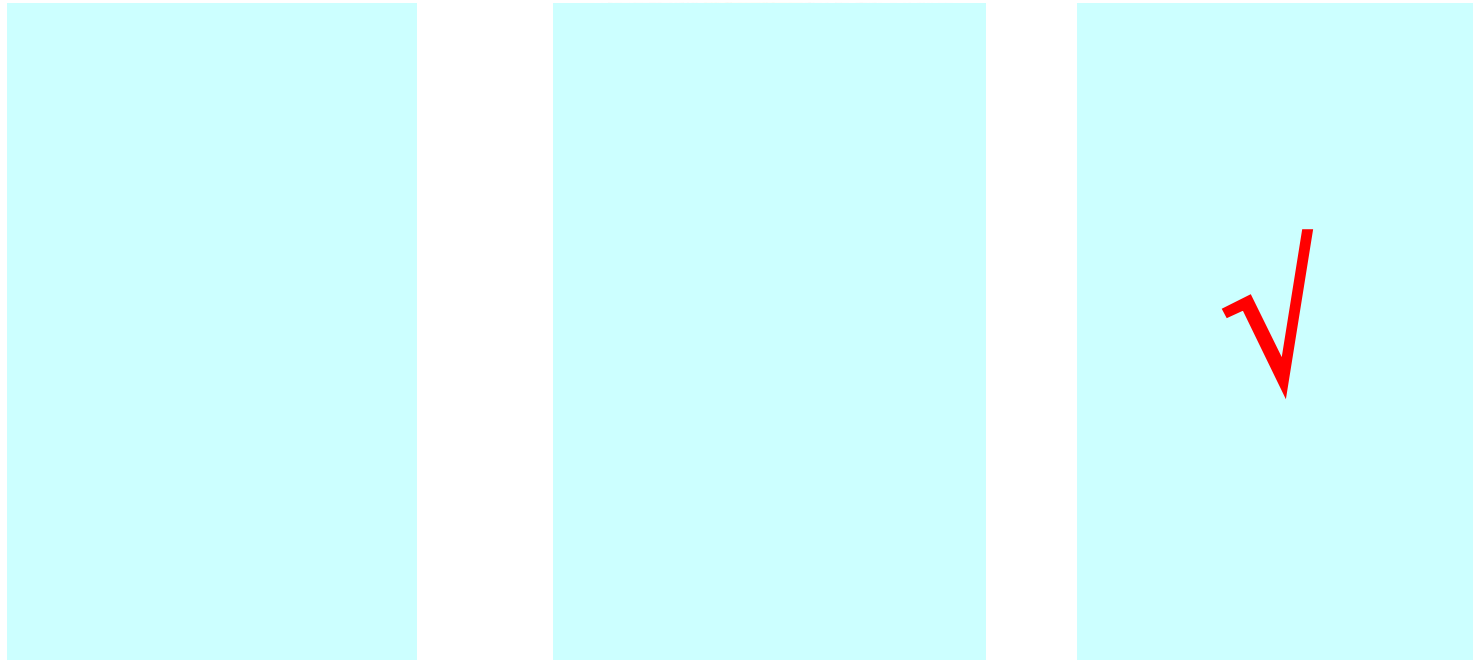


**Place your bet on the location of the King!**



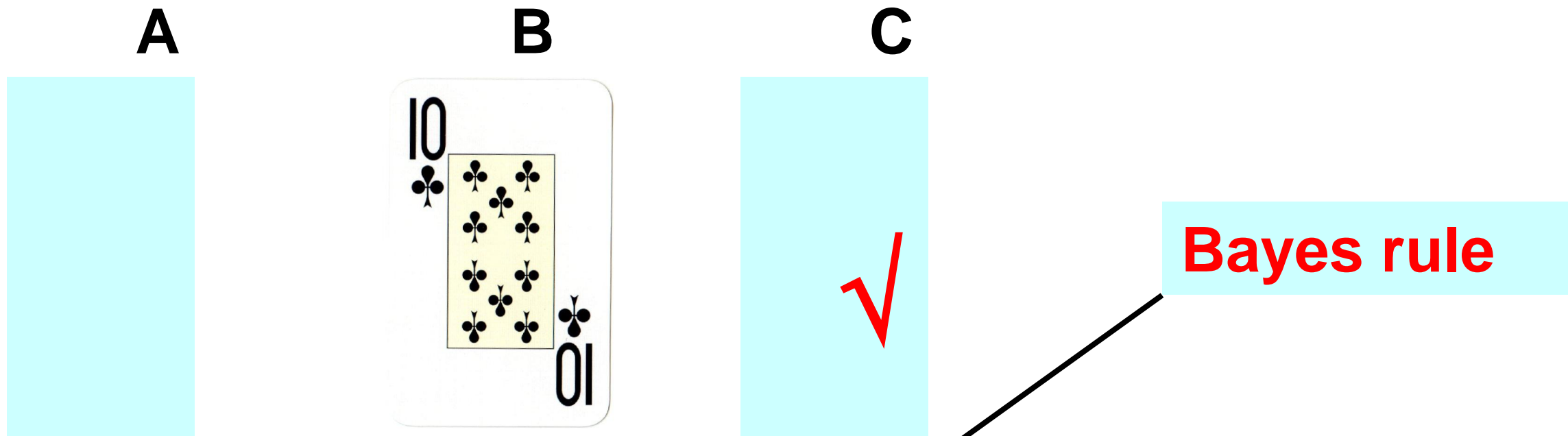
# Using Bayes rule

- Cards game:



**Do you want to  
change your bet?**

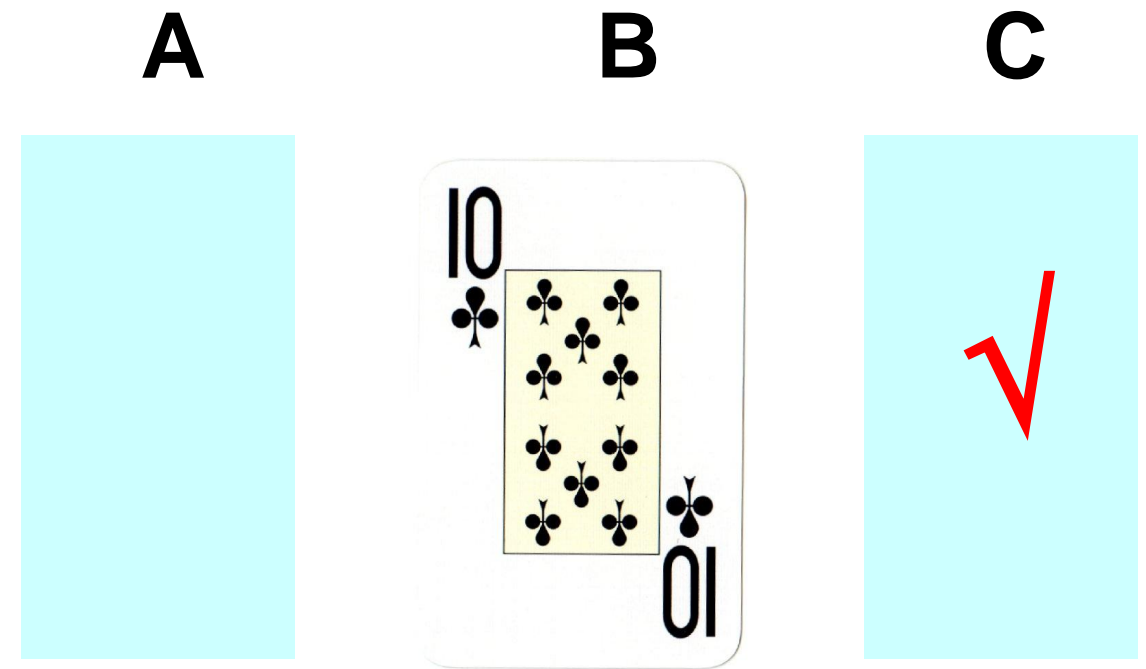
# Using Bayes rule



Computing the (posterior) probability:  $P(C = k \mid \text{sel}B)$

$$P(C = k \mid \text{sel}B) = \frac{P(\text{sel}B \mid C = k)P(C = k)}{P(\text{sel}B)}$$
$$= \frac{P(\text{sel}B \mid C = k)P(C = k)}{P(\text{sel}B \mid C = k)P(C = k) + P(\text{sel}B \mid C = 10)P(C = 10)}$$

# Using Bayes rule



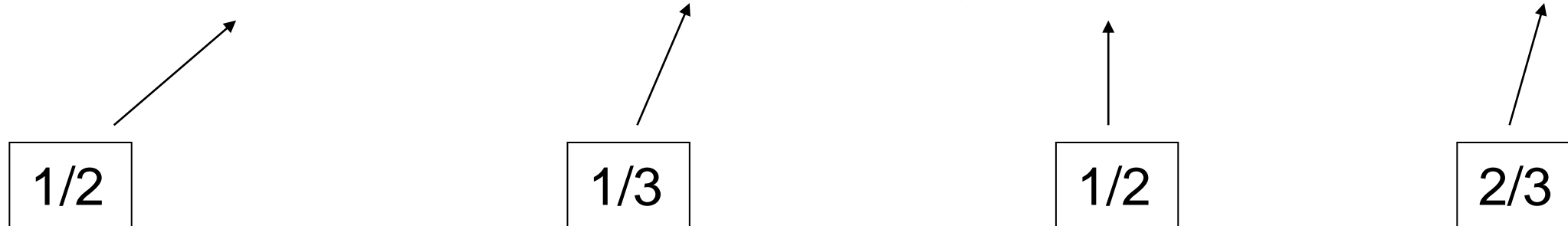
$$P(C=k \mid selB) =$$



$$P(selB \mid C = k)P(C = k)$$

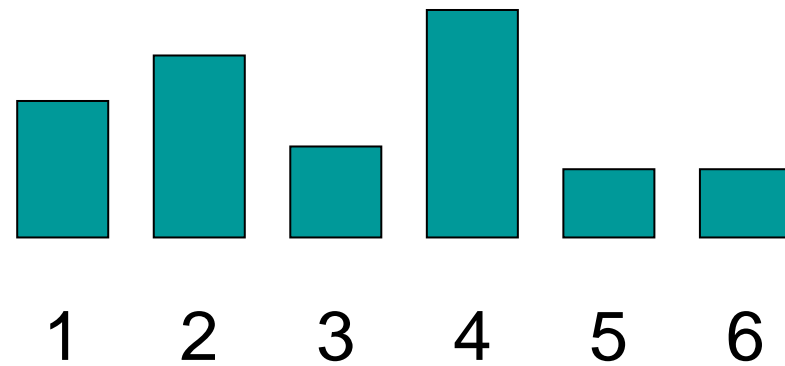
$$= \frac{P(selB \mid C = k)P(C = k) + P(selB \mid C = 10)P(C = 10)}{P(selB \mid C = k)P(C = k) + P(selB \mid C = 10)P(C = 10)}$$

$$= 1/3$$



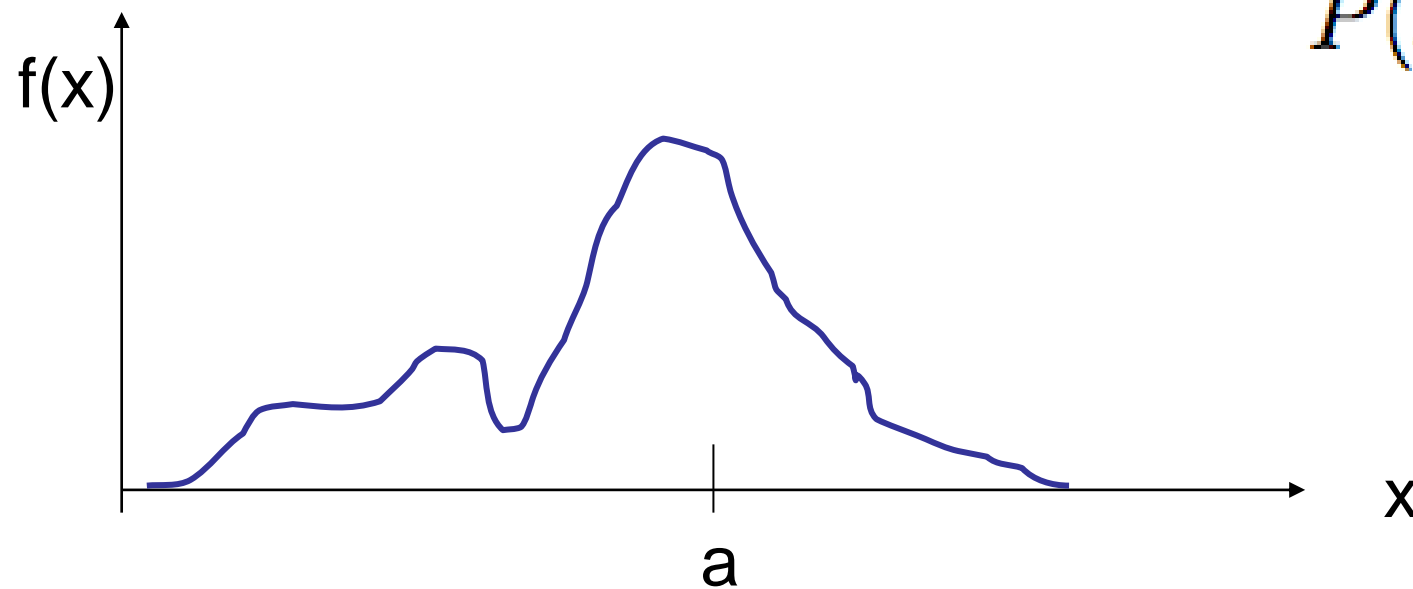
# Probability Density Function

- Discrete distributions



$$\sum_i P(X = x_i) = 1$$

- Continuous: Cumulative Density Function (CDF):  $F(a)$



$$P(x \leq a) = \int_{-\infty}^a f(\tau) d\tau$$

# Cumulative Density Functions

- Total probability

- Probability Density Function (PDF)

$$P(\Omega) = \int_{-\infty}^{\infty} f(x)dx = 1$$

- Properties:

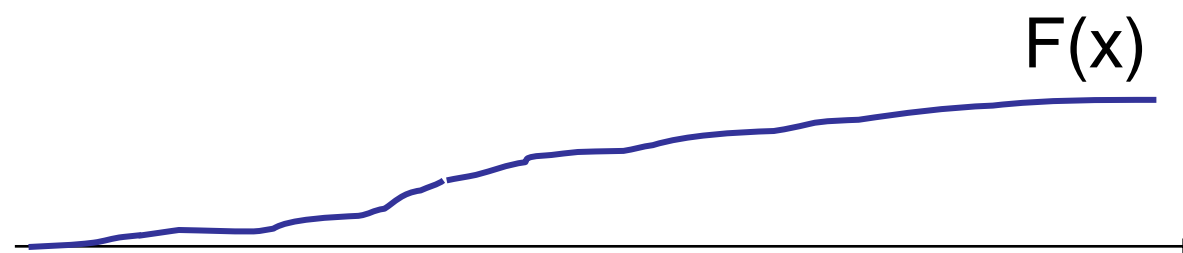
$$\frac{d}{dx}F(x) = f(x)$$

$$P(a \leq x \leq b) = \int_b^a f(x)dx = F(b) - F(a)$$

$$\lim_{x \rightarrow -\infty} F(x) = 0$$

$$\lim_{x \rightarrow \infty} F(x) = 1$$

$$F(a) \geq F(b) \quad \forall a \geq b$$



# Expectations

- Mean/Expected Value:

- Variance:

- Note:

$$E[x] = \bar{x} = \int x f(x) dx$$

- In general:

$$Var(x) = E[(x - \bar{x})^2] = E[x^2] - (\bar{x})^2$$

$$E[x^2] = \int x^2 f(x) dx$$

$$E[g(x)] = \int g(x) f(x) dx$$

# Multivariate

- Joint for (x,y)

- Marginal:

$$P((x, y) \in A) = \int \int_A f(x, y) dx dy$$

- Conditionals:

- Chain rule:

$$f(x) = \int f(x, y) dy$$

$$f(x|y) = \frac{f(x, y)}{f(y)}$$

$$f(x, y) = f(x|y)f(y) = f(y|x)f(x)$$

# Bayes Rule

- Standard form:

- Replacing the bottom:

$$f(x|y) = \frac{f(y|x)f(x)}{f(y)}$$

$$f(x|y) = \frac{f(y|x)f(x)}{\int f(y|x)f(x)dx}$$



# Binomial

- Distribution:

$$x \sim \text{Binomial}(p, n)$$

$$P(x = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

- Mean/Var:

$$E[x] = np$$

$$\text{Var}(x) = np(1 - p)$$

# Uniform

- Anything is equally likely in the region  $[a,b]$
- Distribution:

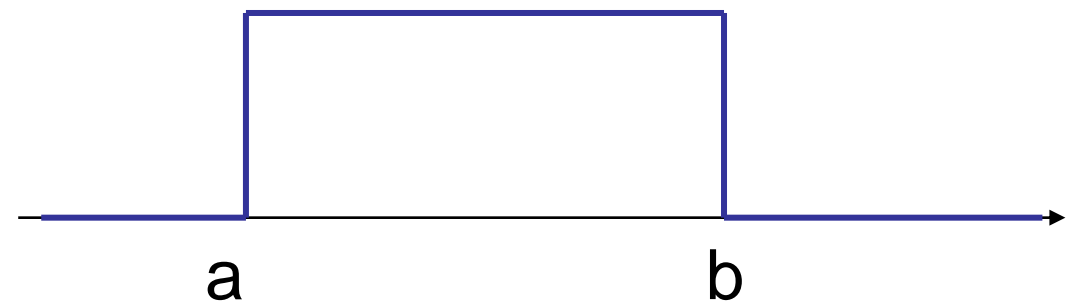
$$x \sim U(a, b)$$

- Mean/Var

$$f(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

$$E[x] = \frac{a+b}{2}$$

$$Var(x) = \frac{a^2 + ab + b^2}{3}$$



# Gaussian (Normal)

- If I look at the height of women in country xx, it will look approximately Gaussian
- Small random noise errors, look Gaussian/Normal

- Distribution:

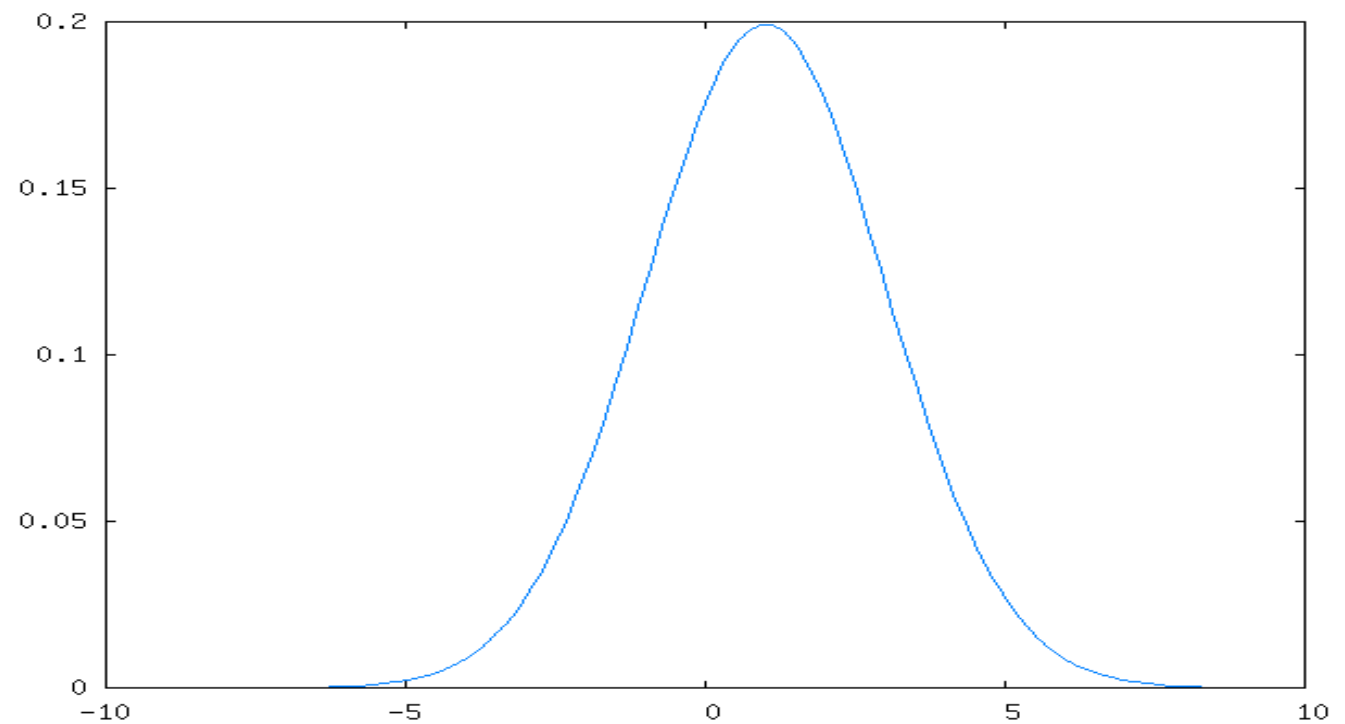
$$x \sim N(\mu, \sigma^2)$$

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- Mean/var

$$E[x] = \mu$$

$$Var(x) = \sigma^2$$



# Why Do People Use Gaussians

- Central Limit Theorem: (loosely)
  - Sum of a large number of IID random variables is approximately Gaussian

# Multivariate Gaussians

- Distribution for vector  $x$

$$x = (x_1, \dots, x_N)^T, \quad x \sim N(\mu, \Sigma)$$

- PDF:

$$f(x) = \frac{1}{(2\pi)^{\frac{N}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)}$$

$$E[x] = \mu = (E[x_1], \dots, E[x_N])^T$$

$$\text{Var}(x) \rightarrow \Sigma = \begin{pmatrix} \text{Var}(x_1) & \text{Cov}(x_1, x_2) & \dots & \text{Cov}(x_1, x_N) \\ \text{Cov}(x_2, x_1) & \text{Var}(x_2) & \dots & \text{Cov}(x_2, x_N) \\ \vdots & & \ddots & \vdots \\ \text{Cov}(x_N, x_1) & \text{Cov}(x_N, x_2) & \dots & \text{Var}(x_N) \end{pmatrix}$$

# Multivariate Gaussians

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{N}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(\mathbf{x}-\mu)^T \Sigma^{-1}(\mathbf{x}-\mu)}$$

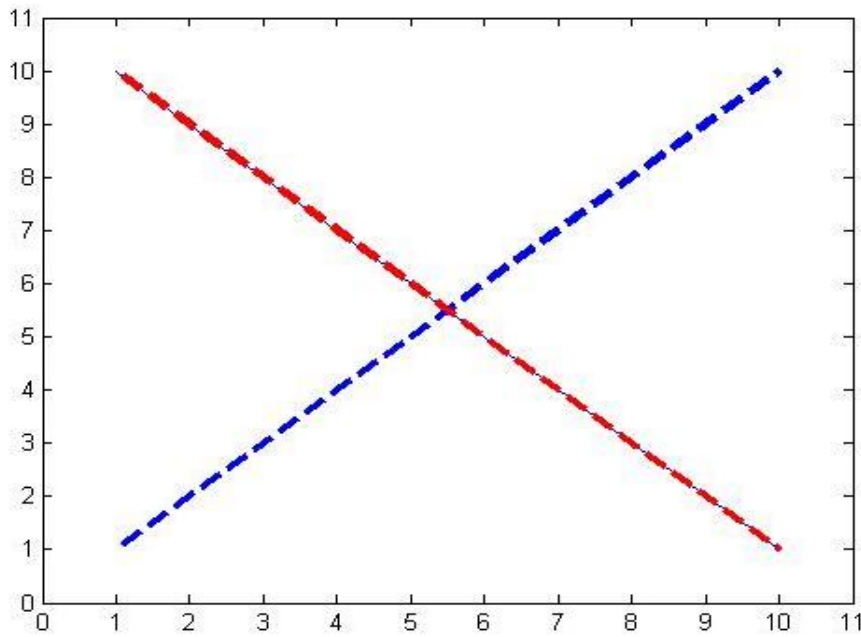
$$E[\mathbf{x}] = \mu = (E[x_1], \dots, E[x_N])^T$$

$$\text{Var}(\mathbf{x}) \rightarrow \Sigma = \begin{pmatrix} \text{Var}(x_1) & \text{Cov}(x_1, x_2) & \dots & \text{Cov}(x_1, x_N) \\ \text{Cov}(x_2, x_1) & \text{Var}(x_2) & \dots & \text{Cov}(x_2, x_N) \\ \vdots & & \ddots & \vdots \\ \text{Cov}(x_N, x_1) & \text{Cov}(x_N, x_2) & \dots & \text{Var}(x_N) \end{pmatrix}$$

$$\text{cov}(x_1, x_2) = \frac{1}{n} \sum_{i=1}^n (x_{1,i} - \mu_1)(x_{2,i} - \mu_2)$$

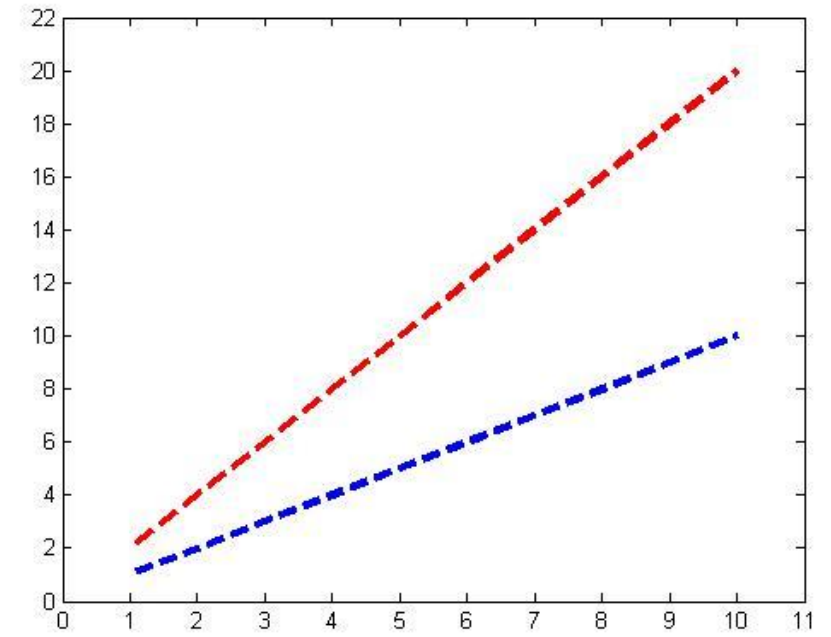
# Covariance examples

Anti-correlated



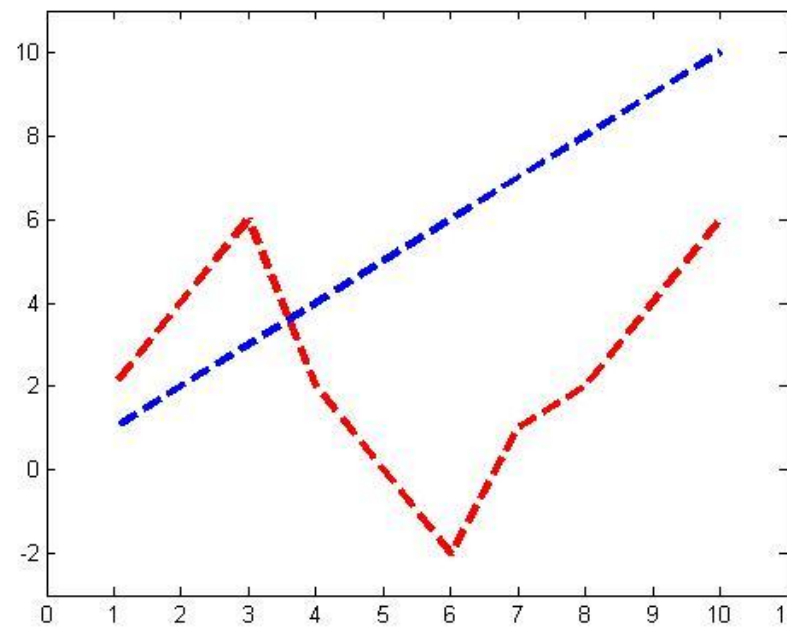
Covariance: -9.2

Correlated



Covariance: 18.33

Independent (almost)



Covariance: 0.6

# Sum of Gaussians

- The sum of two Gaussians is a Gaussian:

$$x \sim N(\mu, \sigma^2) \quad y \sim N(\mu_y, \sigma_y^2)$$

$$ax + b \sim N(a\mu + b, (a\sigma)^2)$$

$$x + y \sim N(\mu + \mu_y, \sigma^2 + \sigma_y^2)$$



# Important points

- Random variables
- Chain rule
- Bayes rule
- Joint distribution, independence, conditional independence