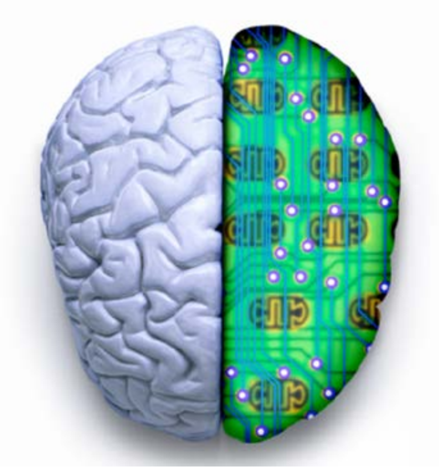
Advanced Artificial Intelligence CM4107 (Week 5)



Probabilistic Models and Inference

Dr. Yann Savoye
School of Computing Science and Digital Media
Robert Gordon University

Module Information

- **Assessment:**

  - Coursework (2 components)
    - Component 1: literature review
    - Component 2: paper implementation

  - No mid-term or final written exam.

    All deadlines are strong:
    - *It will not be possible to upload material after the deadline.*
    - *No deadline extension will be granted. No excuse.*
    - *Only the content submitted via the Moodle will be mark.*

Coursework

- **Submission of the Coursework Part 1:**

  Deadline - Monday, October 30th, 2017 23:00:

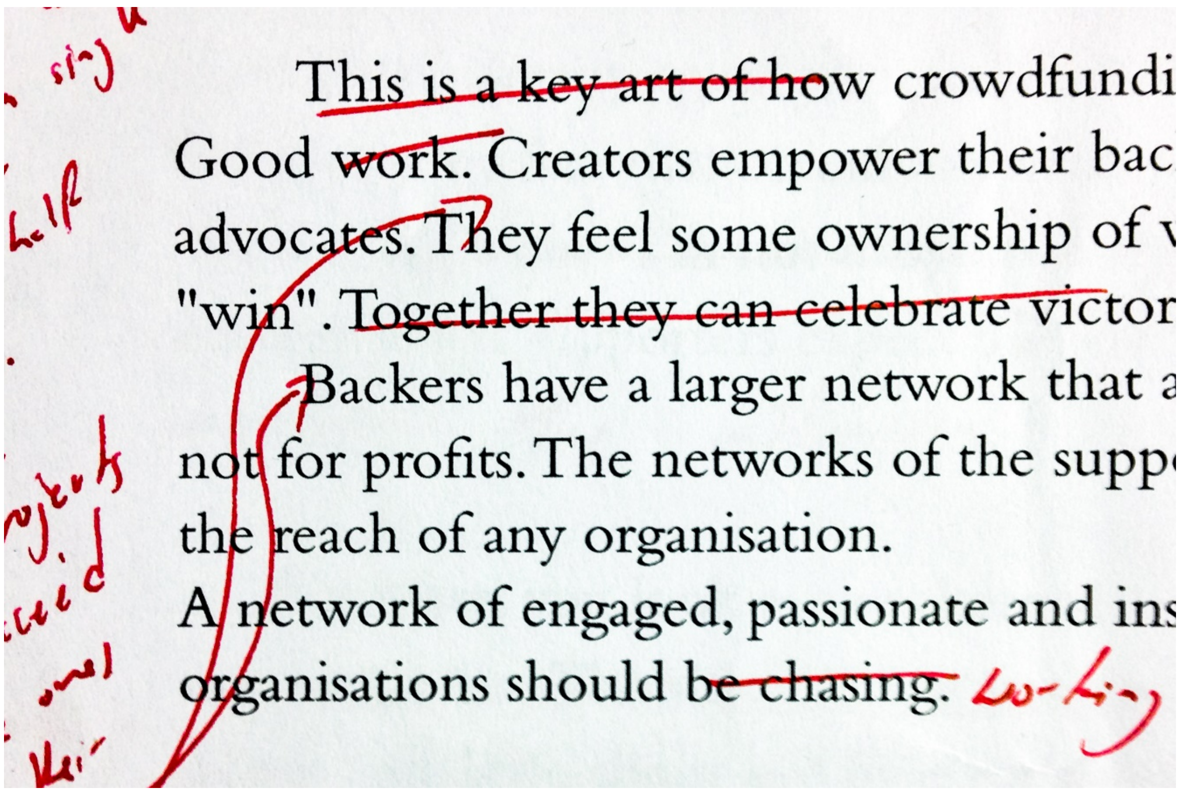
  Activity 1 and Activity 2
    - *2-pages written reports (in PDF format)*
    - *7 slides presentation (in PDF format)*

- You should have finished the first draft and started to rewrite and re-organized the content and how it looks …

Coursework

- **How to write a critical review?**

”Writing is rewriting”



The only kind of writing is rewriting.

Ernest hemingway

Coursework

- **Submission of the Coursework Part 2:**

Deadline - Monday, December 11th, 2017 23:00:

Activity 3 and Activity 4

- *Prolog Programming (code in Prolog)*
- *Paper Implementation (code Java or C++)*

The coursework Part 2 will be released on next Monday….

Overview

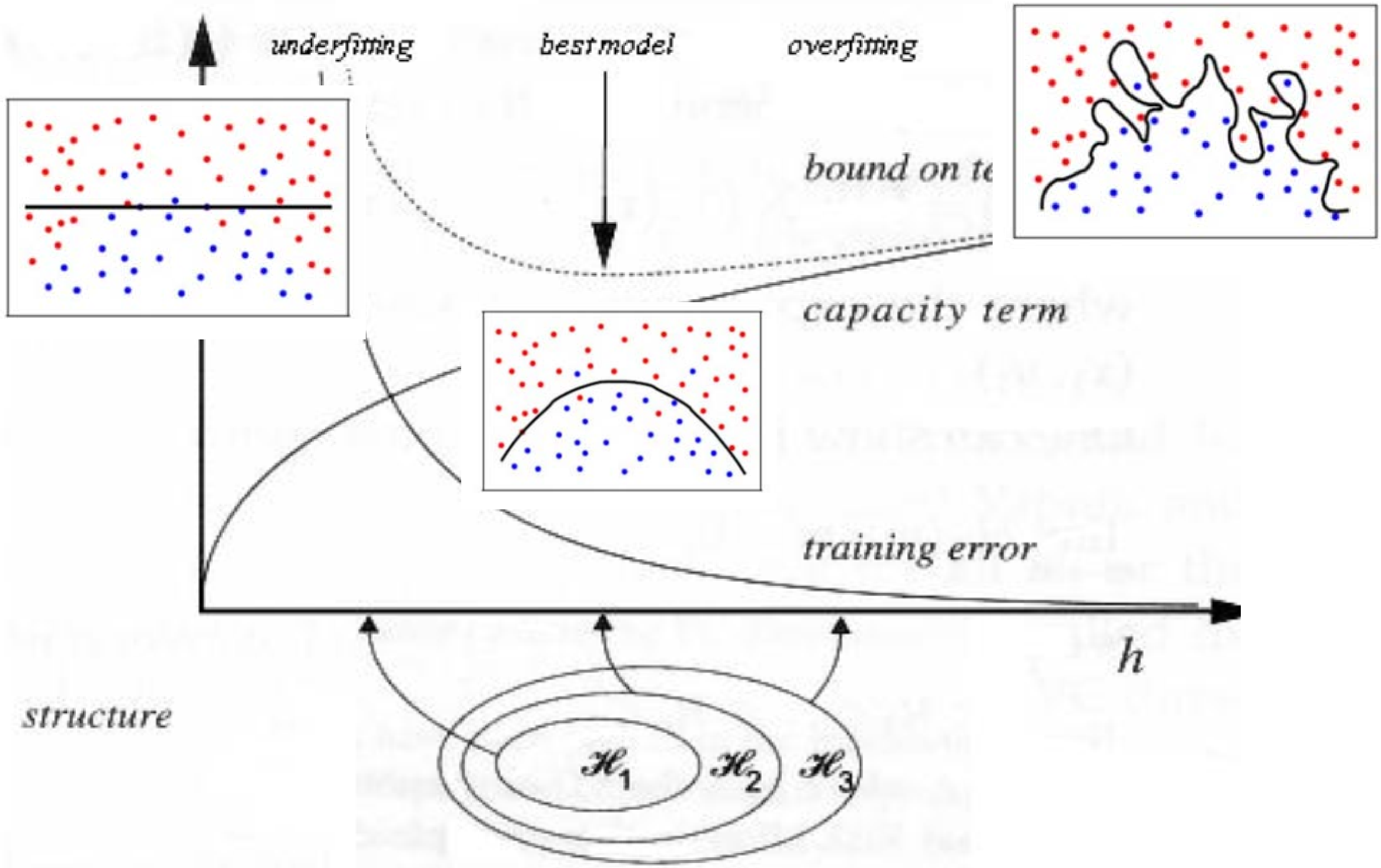
- Part I – Machine Learning and Uncertainty
- Part II – Probability
- Part III – Bayes Reasoning
- Part IV – Bayesian Network
- Part V – Gaussian and Expectation-Maximization

Machine Learning

- Linear Regression
- Support Vector Machine
- Principal Component Analysis
- Random Forest

- **Expectation Maximization**
- **Bayesian Network**
- Neural Network
- Deep Learning

Importance of the Model



*Constructing **models** that predict **data distributions**.*

Uncertainty in AI

- **Uncertain reasoning** is the key feature of a numerous of problems.
- In the 1960s**, medical diagnosis problems** became one of the first attempted application areas of AI programming.
- **No symptom** in medicine **is strictly logically implied by** the existence of any **syndrome.**

- **Representing uncertainty** with **probabilities.**
- **Reasoning** under **uncertainty.**
- We have **to "gamble" rationally.**

Reasoning Under Uncertainty

- **Probabilistic reasoning** uses logic and probability to handle uncertain situations.

- A **probabilistic model** is a mathematical description of an **uncertain situation**.

- **Practical AI systems** shall have to **cope with uncertainty.**

- Three **distinct forms of uncertainty** to cope with:
  1. **Ignorance.** The limits of our knowledge
  2. **Physical randomness** or indeterminism.
  3. **Vagueness.** Some predicates appear to be vague.

Part II – Probabilities and Bayes Theorem

Probabilities

- Use of the **laws of probability** as **coherence constraints** on rational **degrees of belief** (or degrees of confidence).

- The probability of an event refers to the **likelihood** that the event will occur.

- Probability is quantified as a number **between 0 and 1**.
- 0 indicates **impossibility.**
- 1 indicates **certainty.**

- **The higher** the probability of an event, **the more certain** that the event will occur.

A Partition of a Sample Space

- **Sample space:** The set of all possible outcomes.
- **An event:** a collection of possible outcomes.
- We specify a **probability p(x)** for each outcome x such that.

$$p(x) \geq 0, \qquad \sum_{x \in \Omega} p(x) = 1$$

p() = .6

p() = .4

This coin is biased …

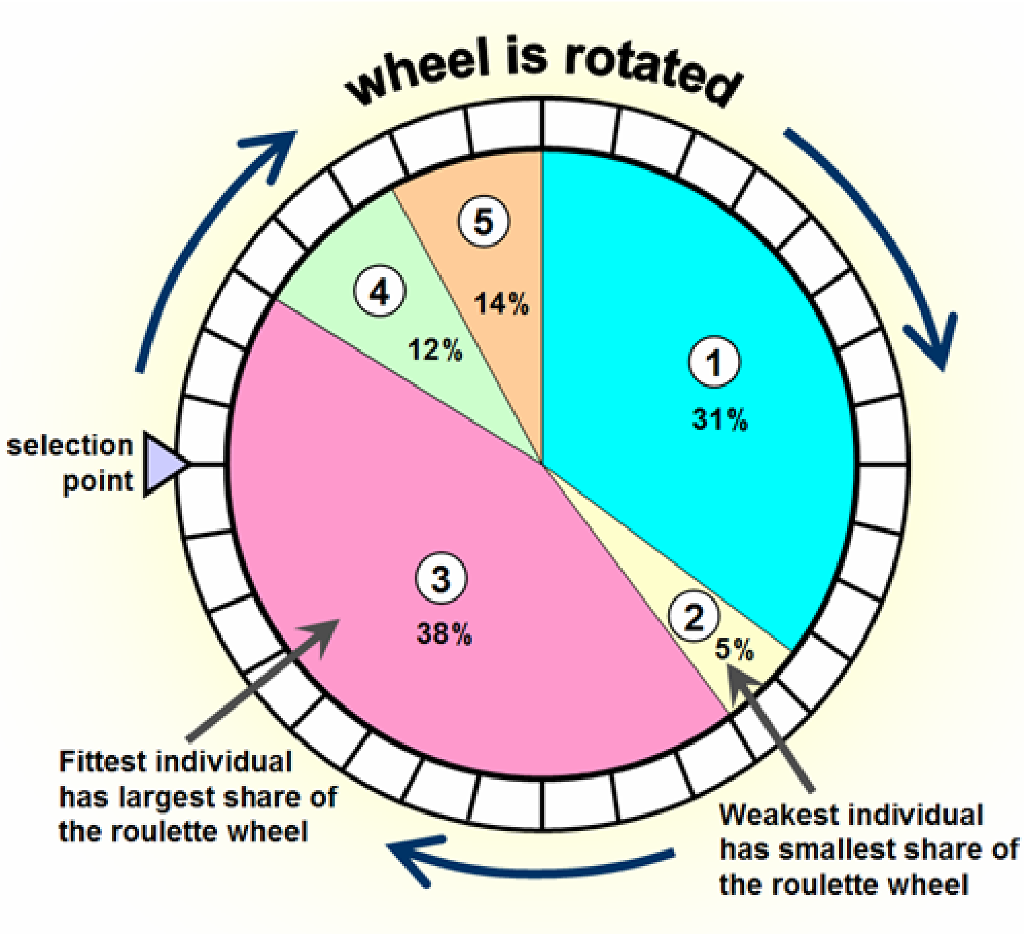
Uniform Random Selection

```
#include <cstdlib>
#include <cstdio>
#include <ctime>

int main()
{
        srand(time(NULL));
        printf("Random numbers:\n");
        float random = 0;
        for (int i = 0; i < 10; ++i) {
                random = (float) rand()/RAND_MAX;
                printf("%f\n", random);
        }
}
```

In an uniform random selection every element is drawn with the ***same probability.***

Weighted Random Selection



*In a weighted random selection, we **select a random item** from a set based on the **importance of its weight**.*

http://www.edc.ncl.ac.uk/highlight/rhjanuary2007g02.php

Part III – Bayesian Reasoning

Bayesian Reasoning

- **Bayesian reasoning** = to reason **probabilistically**
- Stochastic reasoning based on conditional probabilities.

- **Bayesian artificial intelligence** = to produce a thinking system ables to
  - **adapt** to stochastic and changing environments,
  - **recognize** its own limited knowledge
  - **deal** with incomplete evidence.

Four types of **Bayesian reasoning** :

- **causal** - from causes to effects
- **diagnostic** - from effects to causes
- **intercausal** - between causes of a common effect
- **mixed** -combining two or more of the above

Bayes Formula

- The **probability of certain events** are influenced **by our *prior beliefs*** about the **likelihood** of those events.

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

- Think about Bayes rule in terms of **updating our belief** about **a hypothesis *A*** in the light of **new evidence *B*.**

- The power of Bayes' rule: it enables us to compute **P(A|B**) in terms of **P(B|A).**

Bayes Theorem

Posterior probability of the model

Likelihood function of the data

Prior probability of the model

$$P(\text{model} \mid \text{data}) = \frac{P(\text{data} \mid \text{model})\; P(\text{model})}{P(\text{data})}$$

Evidence

Our posterior belief P(A|B) is calculated by multiplying our prior belief P(A) by the likelihood P(B|A) that B will occur if A is true.

Example

Drew Carey

Drew Barrymore

What is the probability of being called "*drew*" given that you are a male?

What is the probability of being a male?

$$P(\text{male} \mid \text{drew}) = \frac{P(\text{drew} \mid \text{male})\ P(\text{male})}{P(\text{drew})}$$

What is the probability of being named "*drew*"?

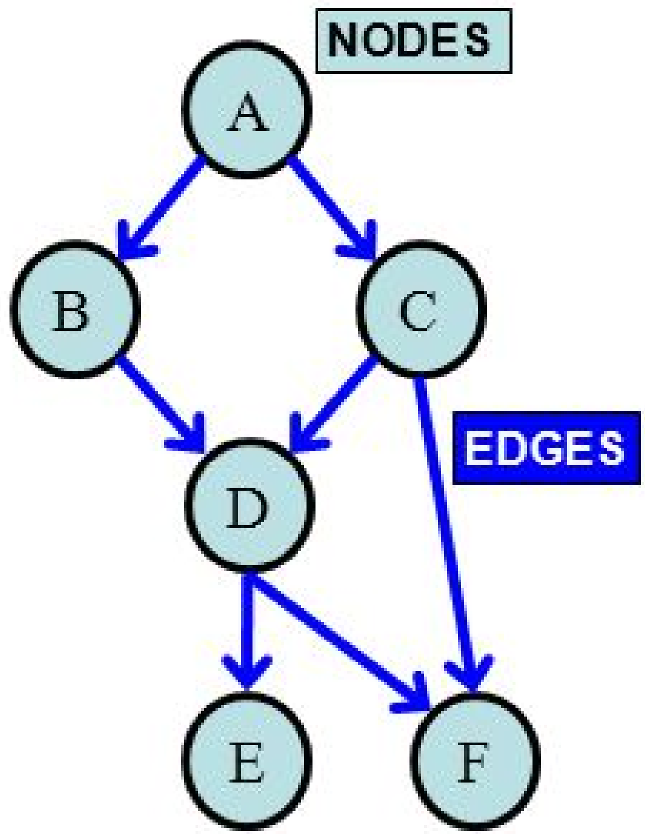
Part IV – Bayesian Network

Bayesian Network

P(B|A)

- A **Bayesian network**: a graph in which variable are nodes.
- Direct **edge** = direct **probabilistic** dependence
- **Graphical representation** of **probabilistic relationship**
- **Explicit representation** of conditional independencies

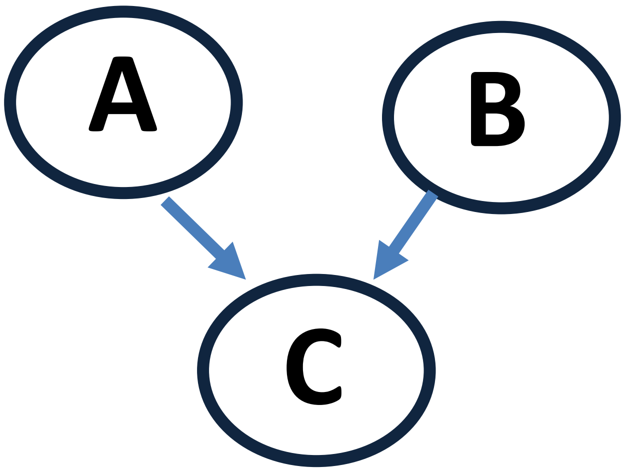
Bayesian Network



- A **Bayesian network** is a probabilistic graphical model that represents a set of random **variables** and their **conditional dependencies**.

- A **Bayesian network** could represent the probabilistic relationships between causes and consequences.

Simple Bayesian Network



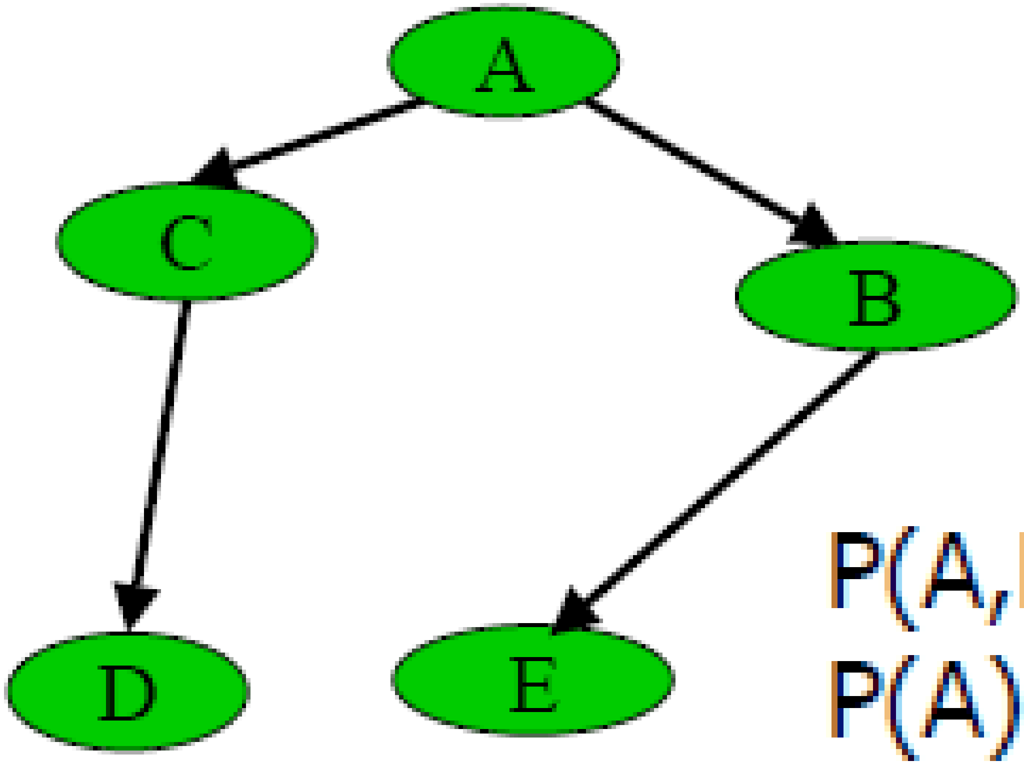
$$p(A,B,C) = p(C|A,B)p(A|B)p(B)$$
$$= p(C|A,B)p(A)p(B)$$

$$P(X_1, X_2, X_3 ..., X_N) = \prod_{i=1}^{n} P(X_i|parents(X_i))$$

The full joint distribution

The graph-structured approximation

Another Example



A tree (with root A)

$P(A,B,C,D,E)=$
$P(A)P(B|A)P(C|A)$
$P(D|C)P(E|B)$

Bayesian Network Application

- Diagnosis:

$$P(\text{cause}|\text{symptom})=?$$

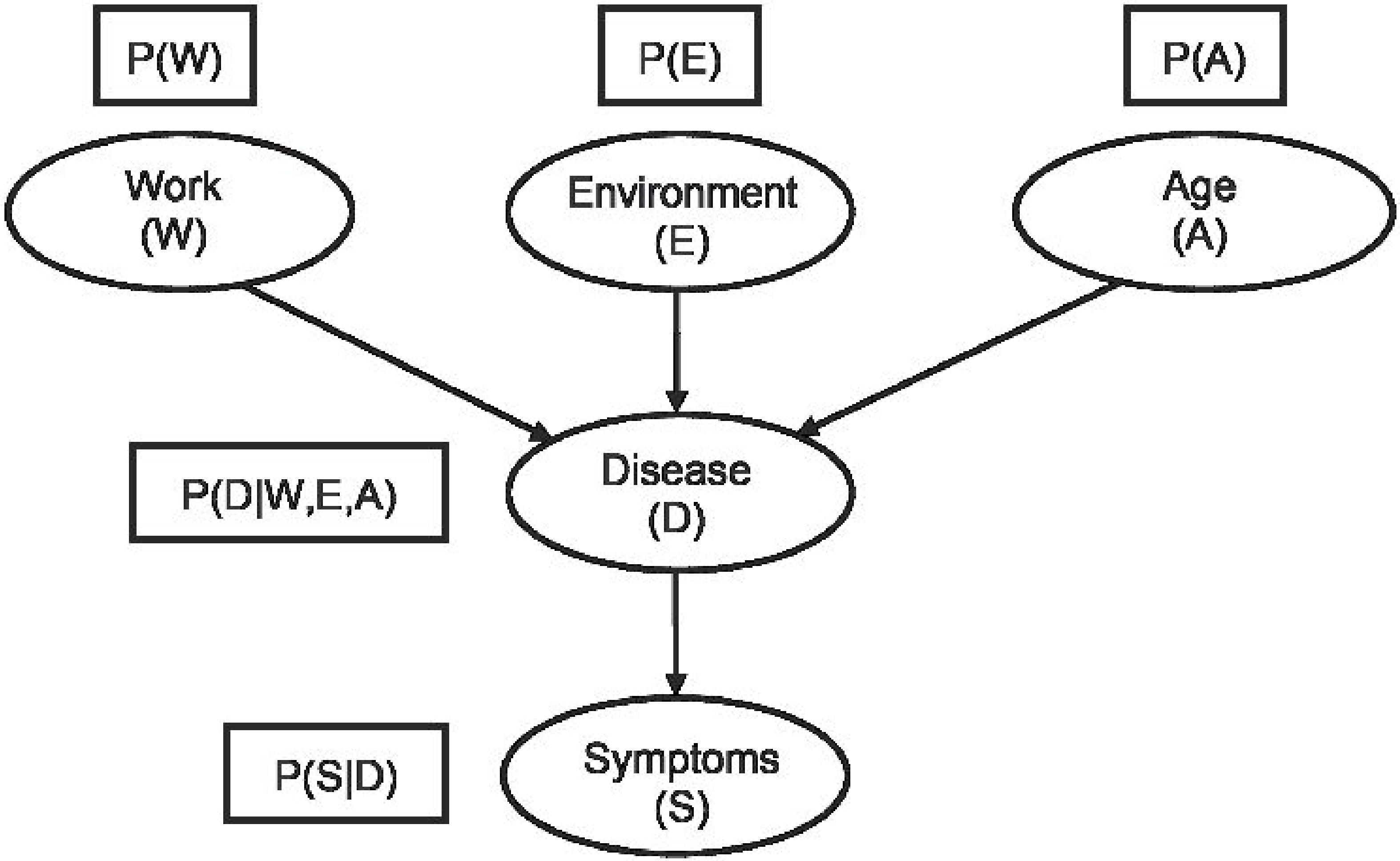
- Prediction:

$$P(\text{symptom}|\text{cause})=?$$

- Classification:

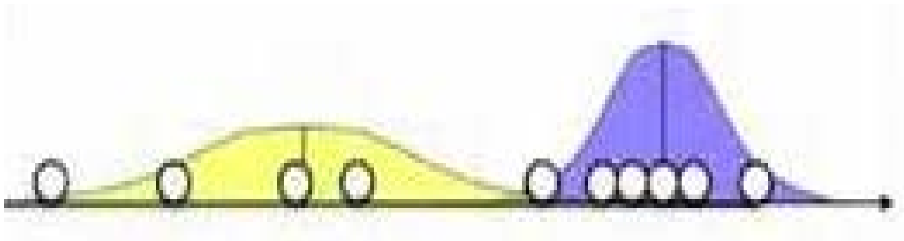
$$\max_{class} P(\text{class}|\text{data})$$

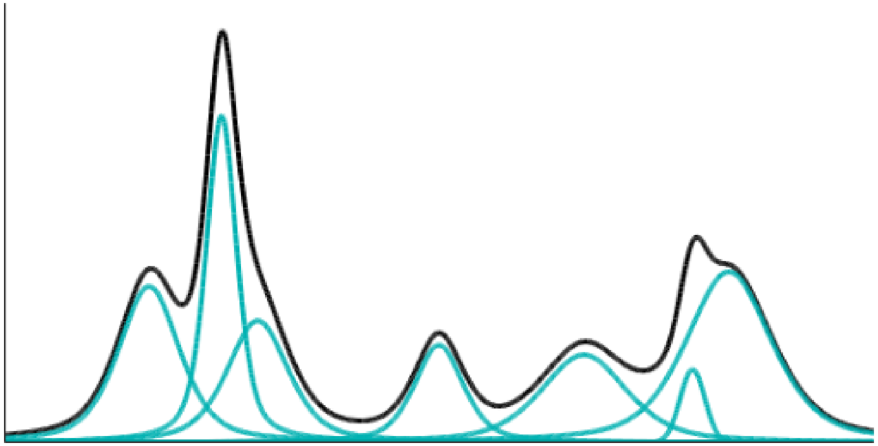
Diagnosis


P(W)
P(E)
P(A)
Work
(W)
Environment
(E)
Age
(A)
P(D|W,E,A)
Disease
(D)
P(S|D)
Symptoms
(S)

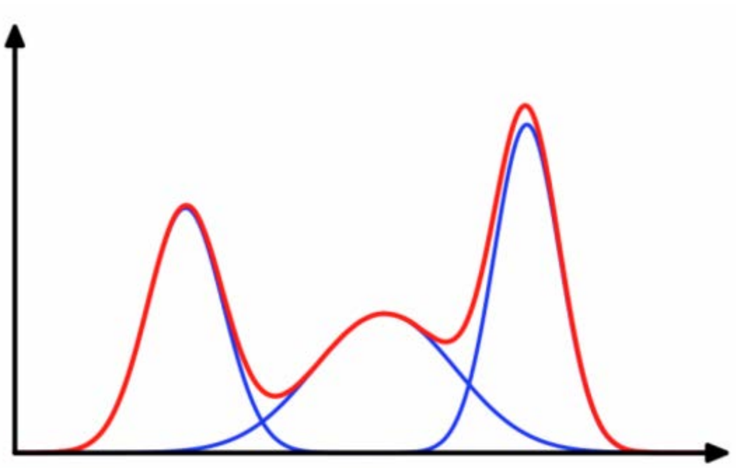
Part V – Gaussian and EM

Gaussian Mixture Model

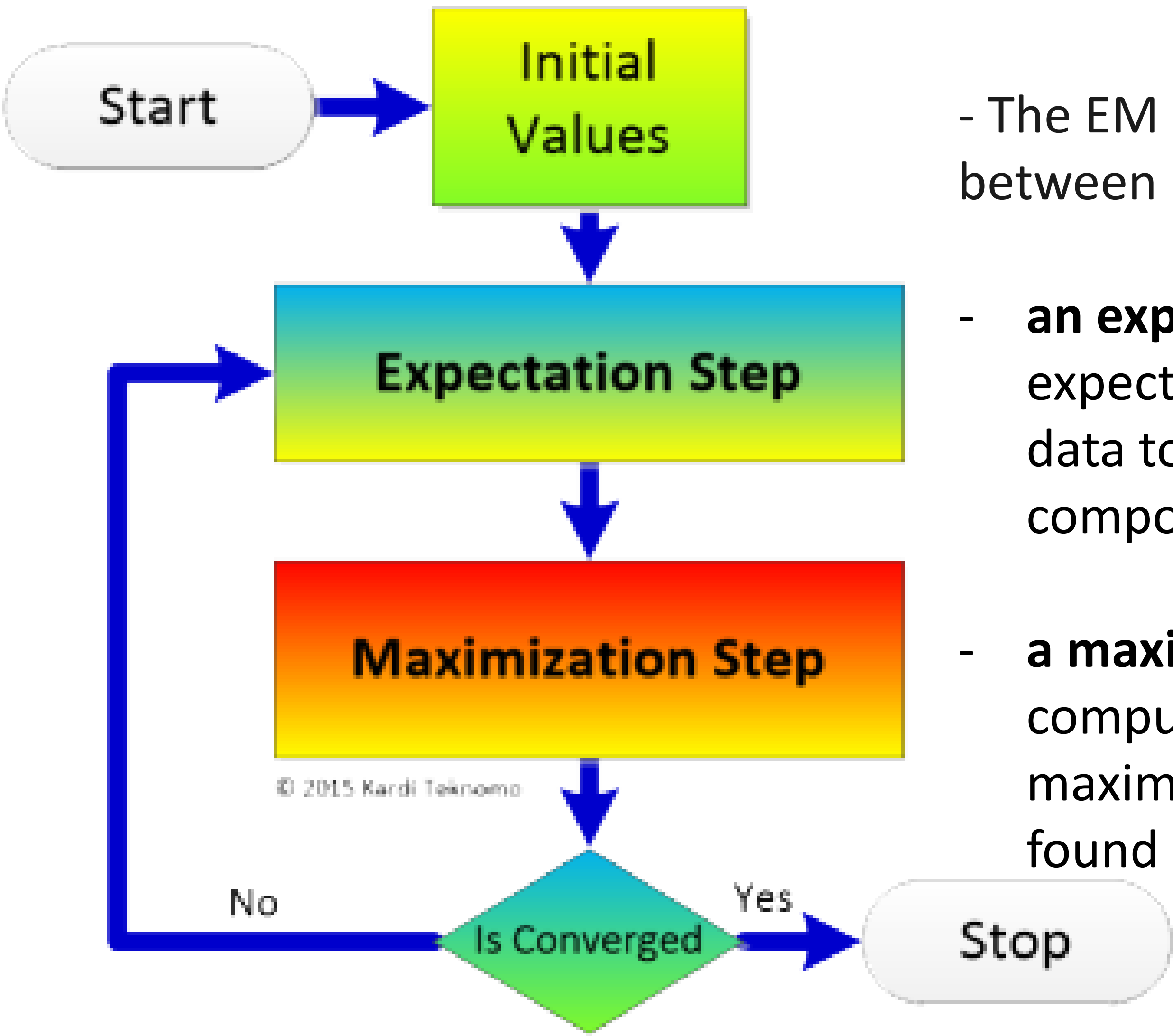




*A **probabilistic model** assuming that all the data points are generated from a **mixture of Gaussian** distributions with unknown parameters.*



Expectation Maximization



- **Find maximum likelihood.**

- The EM iteration alternates between performing:

- **an expectation (E) step**, you get expected probability of each data to belong to a certain component.

- **a maximization (M) step** computes parameters maximizing the expectation found on the *E* step.

Lab Activities

- Activity 1: Weighted Random Sampling (50min)
- Break (10min)
- Activity 2: Bayesian Network (50min)

References

[1] - Worked Examples Total Probability and Bayes Theorem - Alastair Young
[2] - Week 2: Conditional Probability and Bayes formula - Luc Rey-Bellet
[3] - Introduction to Machine Learning - Smola and Vishwanathan
[4] - Doing Bayesian Data Analysis: A Tutorial with R and BUGS - John K. Kruschk
[5] - Some Exercices in Bayesian Inference - Puza and ONeill
[6] - Introduction to Bayesian Learning - Aaron Hertzmann
[7] - Conditional Probability, Independence and Bayes Theorem - Orloff and Bloom
[8] - Introduction to Bayesian methods - David Sontag
[9] - The EM-algorithm - Henrik Hult
[10] - Likelihood and Bayesian Inference - Joe Felsentein
[11] - The EM algorithm - Maneesh Sahani
[12] - Naive-Bayes Classification Algorithm - Cristian Mihaescu
[13] - Bayes theorem and Bayesian inference - Duncan Fyfe Gillies
[14] - Expectation Maximization - Prof. Rosenberg
[15] - Machine Learning - Mitchell and Bar-Joseph
[16] - Detection and Estimation Theory, EM Algorithm - Namrata Vaswani
[17] - Expectation-Maximization - Nuno Vasconcelos
[18] - The Expectation Maximization Algorithm - Sean Borman
[19] - The Naive Bayes Model and the EM Algorithm - Michael Collins
[20] - A Gentle Tutorial of the EM Algorithm - Jeff A. Bilmes

References

[21] - CS 446: Machine Learning - Dan Roth - University of Illinois, Urbana-Champaign

[22] - PCA - Principal Component Analysis - Matthias Scholz, Ph.D. thesis

[23] - Machine Learning / Course Overview & Lecture 1 - Agarwal and Bhattacharya

[24] - Unsupervised learning: Clustering, Ata Kaban, The University of Birmingham

[25] - CS229 Lecture notes - Andrew Ng

[26] - The Expectation-Maximisation (EM) algorithm, Alastair Turl, School of Computer Science

[27] - K-Nearest Neighbours and Instance based learning , Ata Kaban

[28] - Machine Learning / Convnets - Nando de Freitas

[29] - Bayesian Machine Learning - Lecture 1 - Guido Sanguinetti

[30] - Introduction to Machine Learning - Lecture 1 : Overview - Chaohui Wang

[31] - Machine Learning (Extended) - Dr. Ata Kaban

[32] - CS 540 Lecture Notes - C. R. Dyer - Machine Learning (Chapter 18.1 - 18.3)

[33] - The Expectation-Maximisation (EM) algorithm, Alastair Turl

[34] - The EM Algorithm - Ajit Singh

[35] - Bayesian Learning - Emanuel Kitzelmann

[36] - Bayesian AI - Ann E. Nicholson and Kevin B. Korb

[37] - Bayesian networks - Jiri Klema

[38] - Naive Bayes Classifier - Thomas Bayes

[39] - Practice with Bayes theorem - Prof Geoffrey Goodhill

[40] - Total Probability and Bayes Theorem