DM825 (5 ECTS - 4th Quarter) Introduction to Machine Learning

Introduktion til maskinlœring

Marco Chiarandini
adjunkt, IMADA
www.imada.sdu.dk/~marco/

Machine Learning

A computer program is said to **learn** from experience E with respect to some class of tasks T and performance measure P,
if its performance at tasks in T, as measured by P, improves with experience E.

Tom M. Mitchell (1997) Machine Learning p.2

Machine Learning

A computer program is said to **learn** from experience E with respect to some class of tasks T and performance measure P,
if its performance at tasks in T, as measured by P, improves with experience E.

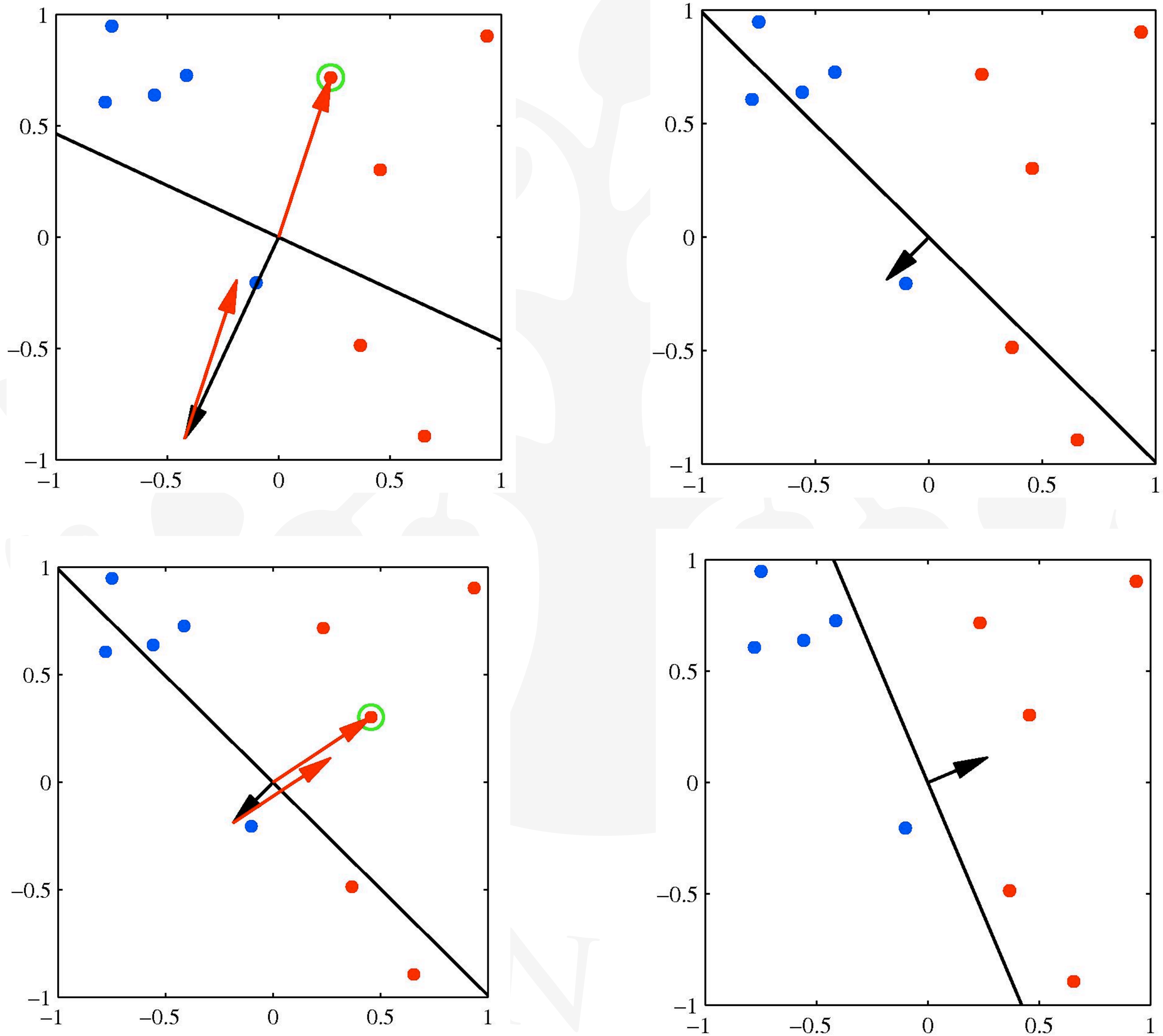
Tom M. Mitchell (1997) Machine Learning p.2

Core objective of a learner: generalize from its experience.

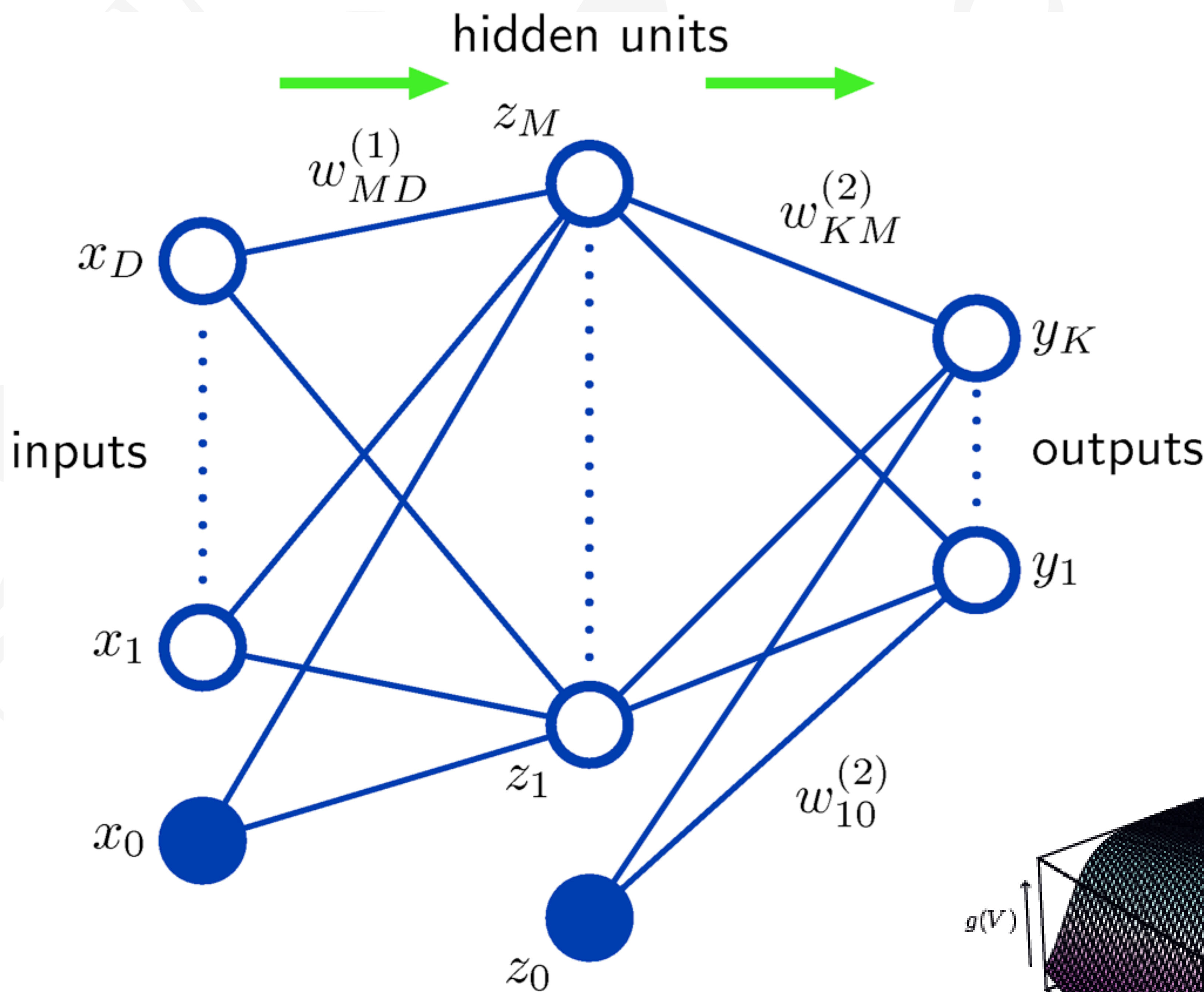
Training examples from experience come from unknown probability distribution. The learner has to extract something to produce a useful answer in new cases.

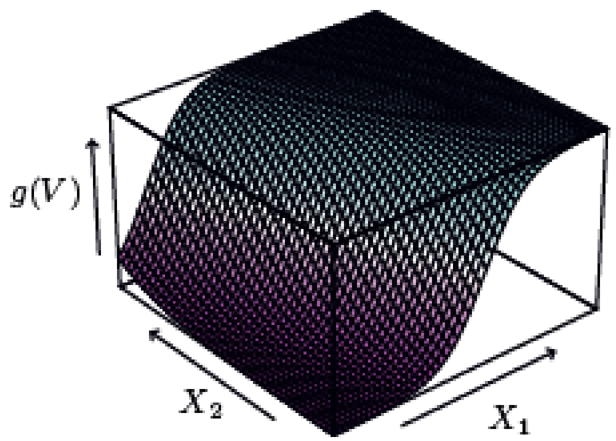
Contents

Perceptron algorithm

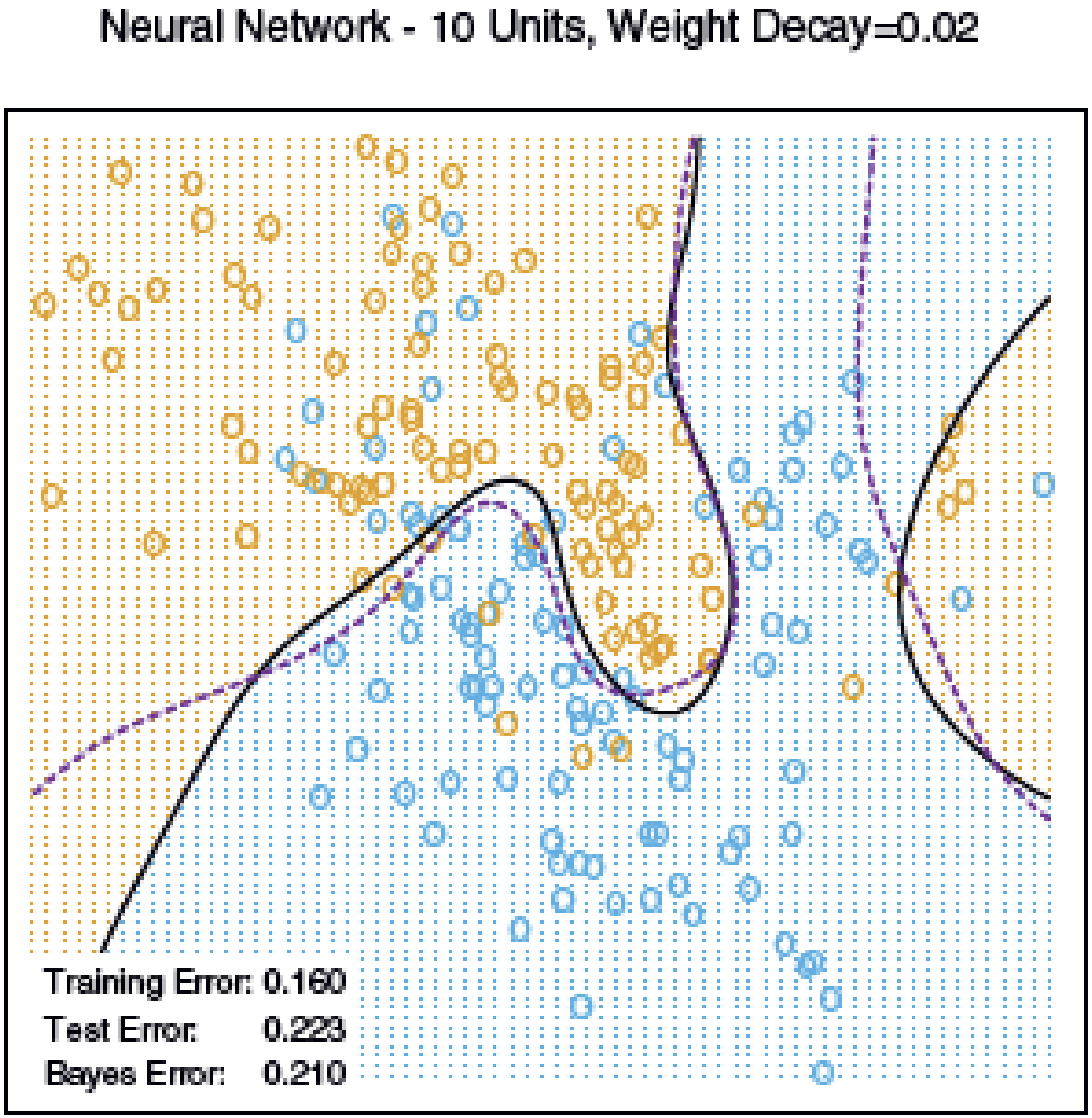


Multilayered Neural Networks



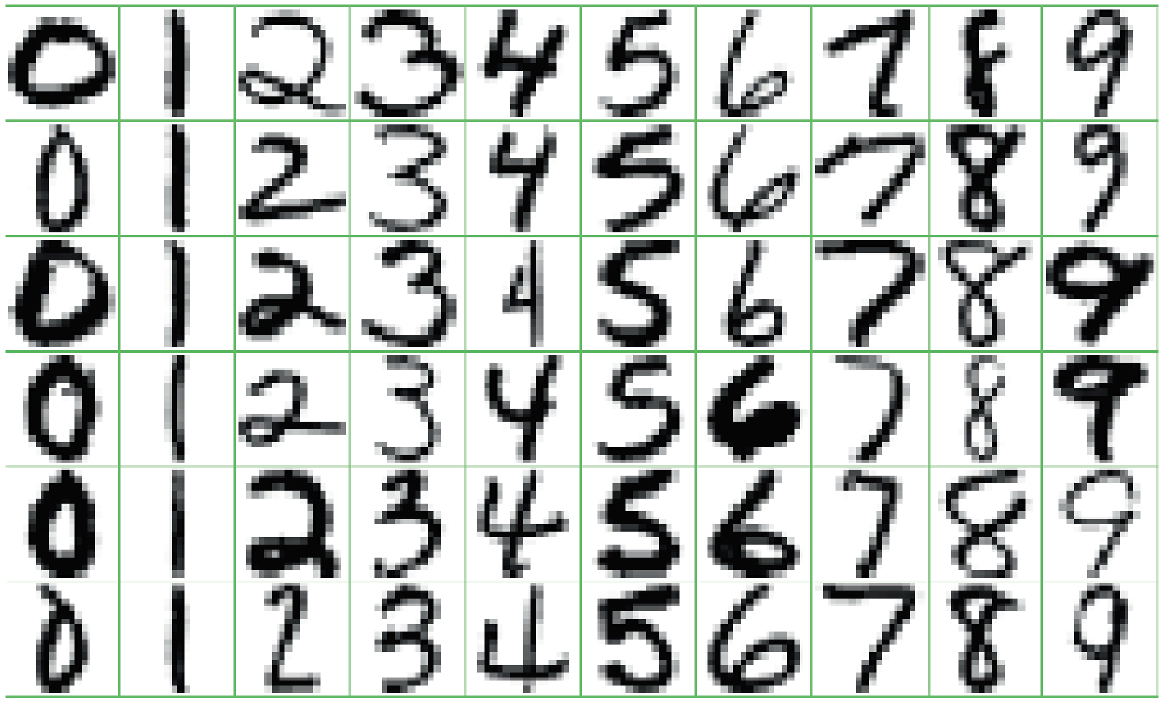


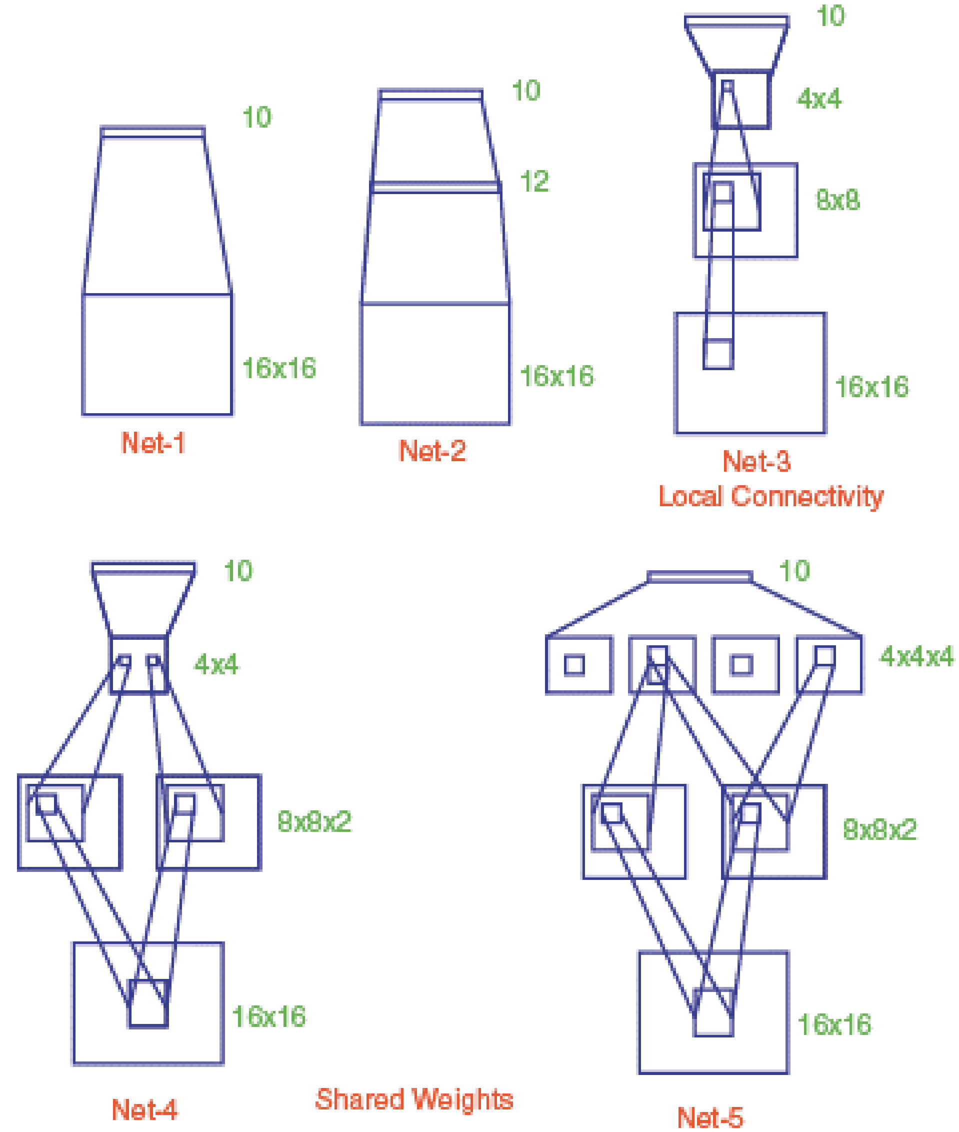
Applications



Applications

Handwritten digit recognition





Humans are at 0.2% – 2.5 % error
400–300–10 unit MLP = 1.6% error
LeNet: 768–192–30–10 unit MLP = 0.9% error

Graphical Models

Allow to represent our prior knoweldge
and to use a general suite of algorithms to make inference and to improve our models for a specific application domain

Complex systems involve uncertainty => Probability framework

interralated aspects of the system are modelled as random variables

Example: Medical diagnosis

- two deases: Fly and Hayfever
- they are not mutually exclusive
- Season might be correlated with them
- symptoms such as Congestion and Muscle Pain

4 random variables:
Flu = {true,false}; Hayfever = {true, false}
Season = {fall, winter, spring, summer}
Congestion = {true, false}
MusclePain = {true, false}

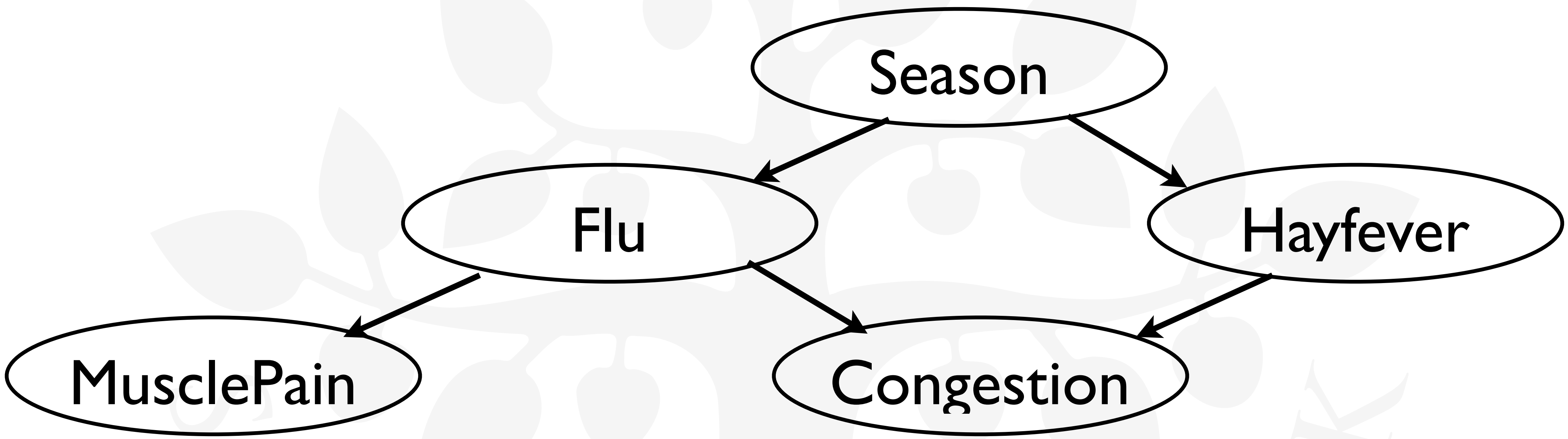
2x2x4x2x2=64
possible prob. values
for joint distribution

P(Flu=true | Season=fall, Congestion=true, MusclePain=false)

If the number of variables grows the problem becomes intractable

Example: continued

Graphical models use graph-based representation to encode independencies



F and H independent given Season
C and S independent given F and H
M and H,C independent given F
M and C independent gien F

We thus only need to define
3+ 4 +4 +4 +2 =17 parameteers

$$P(S,F,H,C,M)=P(S)P(F|S)P(H|S)P(C|F,H)P(M|F)$$

Bayesian Learning

What can we do from here?

- Inference: Complexity issues O(2^n)
- Learning (parameters and structure)

Bayesian Learning

What can we do from here?

- Inference: Complexity issues O(2^n)
- Learning (parameters and structure)

Thumbtack Experiment

Bayesian Learning

What can we do from here?

- Inference: Complexity issues O(2^n)
- Learning (parameters and structure)

Thumbtack Experiment

Flip the thumbtack in the air and observe the number of times it lands with head and tail

We wish to learn how much the probability deviates from 0.5

Bayesian Learning

What can we do from here?

- Inference: Complexity issues O(2^n)
- Learning (parameters and structure)

Thumbtack Experiment

Flip the thumbtack in the air and observe the number of times it lands with head and tail

We wish to learn how much the probability deviates from 0.5

Bayesian Learning

What can we do from here?

- Inference: Complexity issues O(2^n)
- Learning (parameters and structure)

Thumbtack Experiment

Flip the thumbtack in the air and observe the number of times it lands with head and tail

We wish to learn how much the probability deviates from 0.5

Suppose we observe 3 heads in 10 tosses.

- With no prior knowledge we would set p=3/10=0.33
- With a prior of 10 heads over 20 tosses we would set p=(3+10)/(10+20)=13/30=0.43
- However if we obtain more data the effect diminshes: (300+1)/1000+2=0.3 and (300+10)/(1000+20)=0.3

Course Organization

Prerequisites

- ✓ MM501 Calculus I
- ✓ MM505 Linear Algebra
- ✓ Basics of Probability Calculus

Final Assessment (5 ECTS)

- Mandatory assignments, pass/fail, internal evaluation by the teacher. Include programming work in R
- 3 hours written exam, Danish 7 mark scale
- External examiner

Course Material

- Text book
  - C.M. Bishop. Pattern recognition and Machine Learning Springer, 2006
  - Slides
- Source code and data sets
- www.imada.sdu.dk/~marco/DM825