Sequential Data

Giampiero Salvi

KTH/CSC/TMH giampi@kth.se

DD2431 HT2017

Independence Assumption

$$\mathcal{D} = \{x_1, x_2, \ldots, x_N\}$$

Assumption: Independent and Identically Distributed (i.i.d.):

$$P(\mathcal{D}) = P(x_1, x_2, \ldots, x_N) = P(x_1)P(x_2)\cdots P(x_N)$$

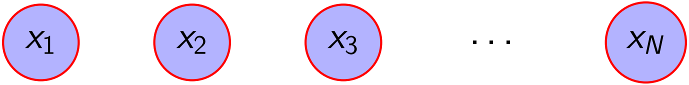
Independence Assumption

$$\mathcal{D} = \{x_1, x_2, \ldots, x_N\}$$

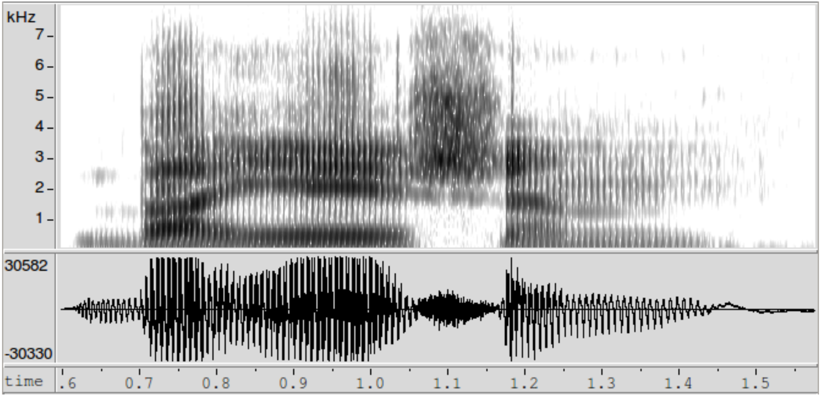
Assumption: Independent and Identically Distributed (i.i.d.):

$$P(\mathcal{D}) = P(x_1, x_2, \ldots, x_N) = P(x_1)P(x_2)\cdots P(x_N)$$

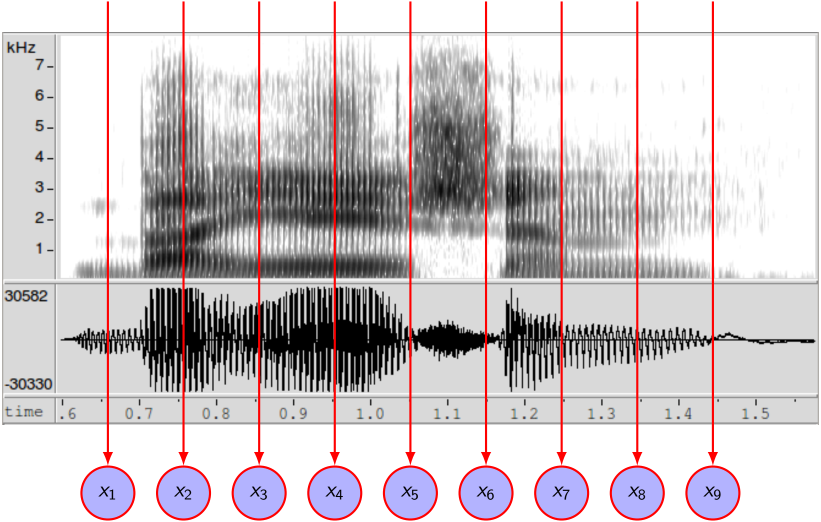
the order of observations is irrelevant!



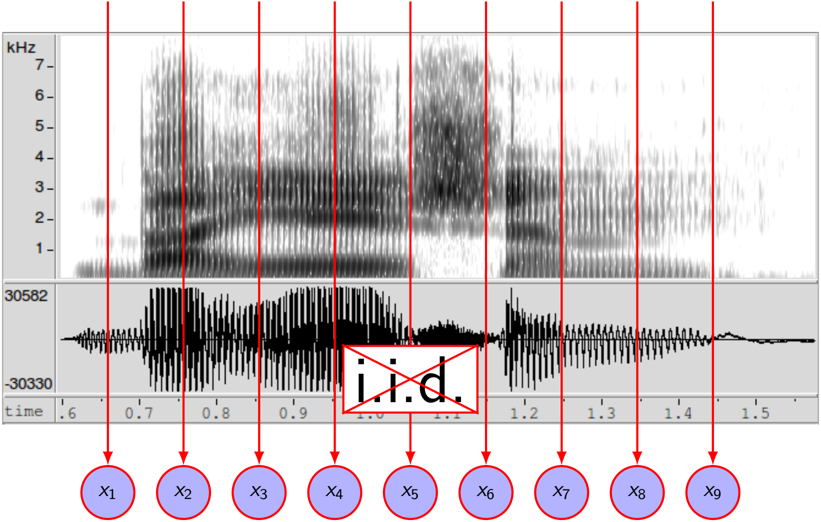
Speech Example



Speech Example



Speech Example

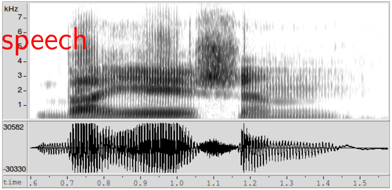


Sequential Data

Time sequences

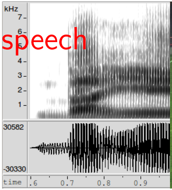
Sequential Data

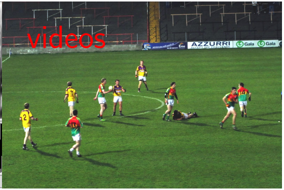
Time sequences



Sequential Data

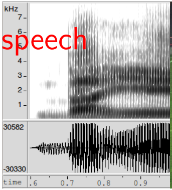
Time sequences





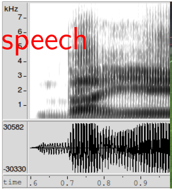
Sequential Data

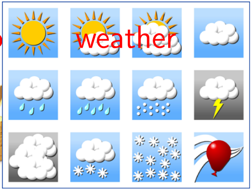
Time sequences



Sequential Data

Time sequences





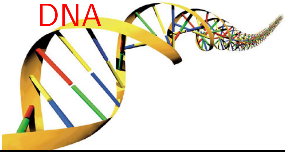
Sequential Data

Time sequences

Sequential Data

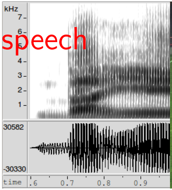
Time sequences

Timeless sequences



Sequential Data

Time sequences

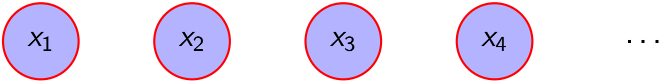


Timeless sequences

text

Nel mezzo del cammin di nostra vita
mi ritrovai per una selva oscura,
ché la diritta via era smarrita.
Ahi quanto a dir qual era è cosa dura
esta selva selvaggia e aspra e forte
che nel pensier rinova la paura!
Tant' è amara che poco è più morte;
ma per trattar del ben ch'i' vi trovai,
dirò de l'altre cose ch'i' v'ho scorte.
Io non so ben ridir com' i' v'intrai,

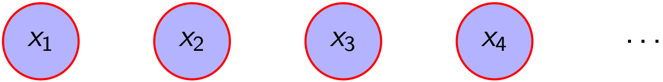
Modelling Sequences with (D)BNs



independence assumption (e.g. i.i.d) not satisfactory

we cannot write $p(x_1, \ldots, x_N) = p(x_1)p(x_2)\cdots p(x_N)$

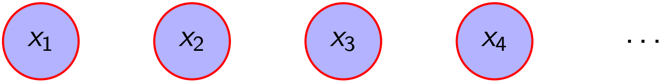
Modelling Sequences with (D)BNs



Most general case, applying product rule recursively $(p(a, b) = p(a)p(b|a))$

$$p(x_1, \dots, x_N) = p(x_1)p(x_2, \dots, x_N|x_1)$$

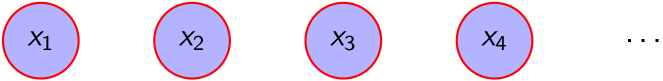
Modelling Sequences with (D)BNs



Most general case, applying product rule recursively ($p(a, b) = p(a)p(b|a)$)

$$p(x_1, \dots, x_N) = p(x_1)p(x_2|x_1)p(x_3, \dots, x_N|x_1, x_2)$$

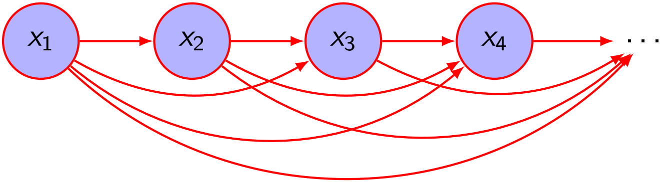
Modelling Sequences with (D)BNs



Most general case, applying product rule recursively $(p(a, b) = p(a)p(b|a))$

$$\begin{aligned} p(x_1, \ldots, x_N) =& p(x_1)p(x_2|x_1)p(x_3|x_1, x_2)\cdots \\ & \cdots p(x_N|x_1, \ldots, x_{N-1}) \end{aligned}$$

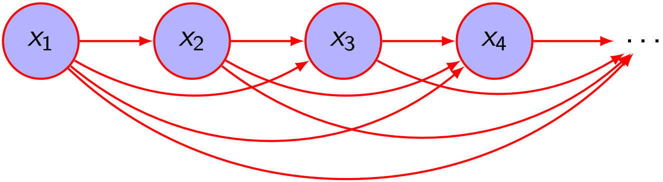
Modelling Sequences with (D)BNs



Most general case, applying product rule recursively $(p(a, b) = p(a)p(b|a))$

$$p(x_1, \ldots, x_N) = p(x_1)p(x_2|x_1)p(x_3|x_1, x_2) \cdots \cdots p(x_N|x_1, \ldots, x_{N-1})$$

Modelling Sequences with (D)BNs



Most general case, applying product rule recursively ($p(a, b) = p(a)p(b|a)$)

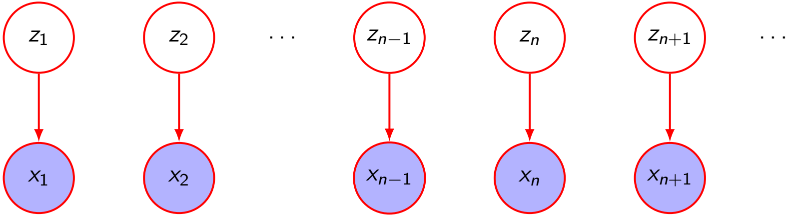
$$p(x_1, \dots, x_N) = p(x_1)p(x_2|x_1)p(x_3|x_1, x_2)\cdots \\ \cdots p(x_N|x_1, \dots, x_{N-1})$$

Grows quadratically with sequence length (N)!!!

State Space Models

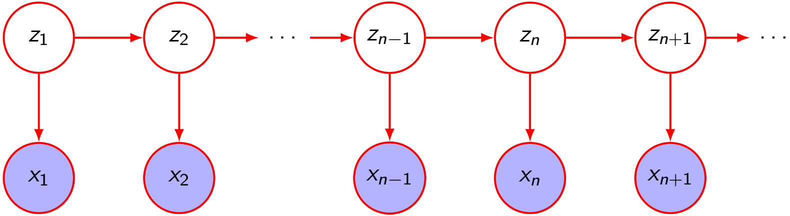
Adding latent variables z_n

Similar to Mixture Model



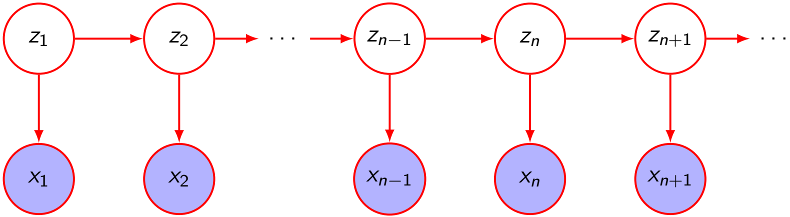
State Space Models

Adding latent variables z_n



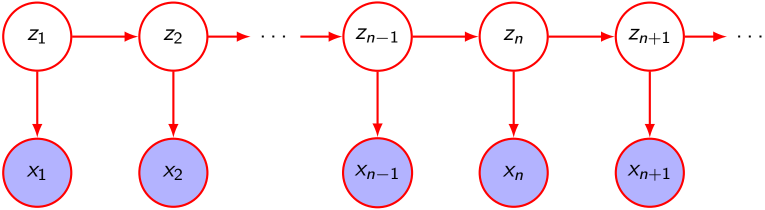
State Space Models

Adding latent variables z_n



We can model indefinitely long dependencies with a limited set of parameters!

State Space Models Instances



- if z_n are continuous: Linear Dynamical Systems
- if z_n are discrete: Hidden Markov Models

Example of continuous z_n: tracking

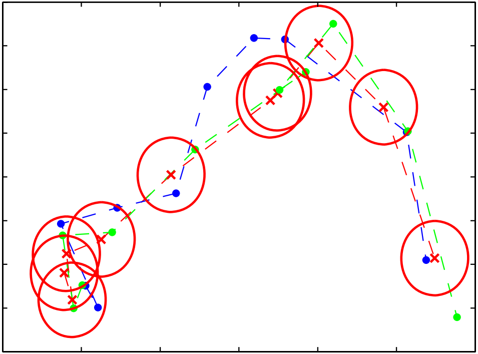
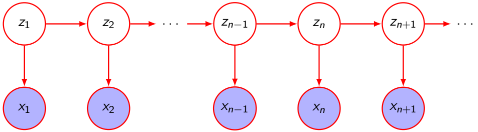


Figure from Bishop

true positions
noisy measurements
inferred posterior (means and covariances)

Example of discrete z_n: speech recognition



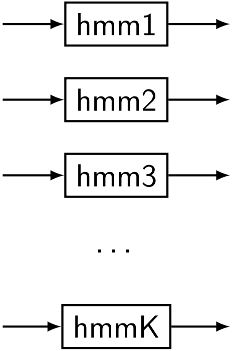
- z_n describes the phonetic class at step n
- x_n describes the acoustic feature vector at step n

Possible inference:

- find likelihood of x_n (isolated word recognition)
- find posterior for z_n (phoneme recognition, continuous speech recognition)
- find best model parameters given sequences of observations (training)

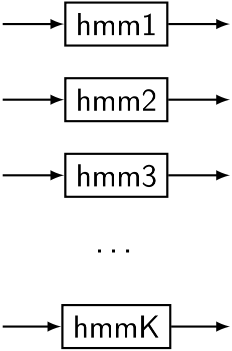
Likelihood and Classification

Example: Isolated Words Recognition (unknown utterance X)



Likelihood and Classification

Example: Isolated Words Recognition (unknown utterance X)



$$P(X|\theta_1)$$

$$P(X|\theta_2)$$

$$P(X|\theta_3)$$

$$\dots$$

$$P(X|\theta_K)$$

Likelihood and Classification

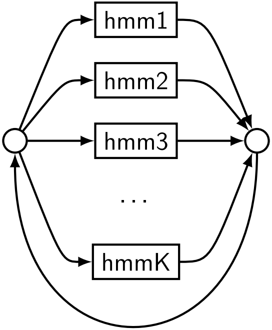
Example: Isolated Words Recognition (unknown utterance X)

$\rightarrow \boxed{\text{hmm1}} \rightarrow$	$P(X\|\theta_1)$
$\rightarrow \boxed{\text{hmm2}} \rightarrow$	$P(X\|\theta_2)$
$\rightarrow \boxed{\text{hmm3}} \rightarrow$	$P(X\|\theta_3)$
$\dots$	$\dots$
$\rightarrow \boxed{\text{hmmK}} \rightarrow$	$P(X\|\theta_K)$

Compare Likelihoods

HMM Inference: Best Path

Example Continuous Speech Recognition

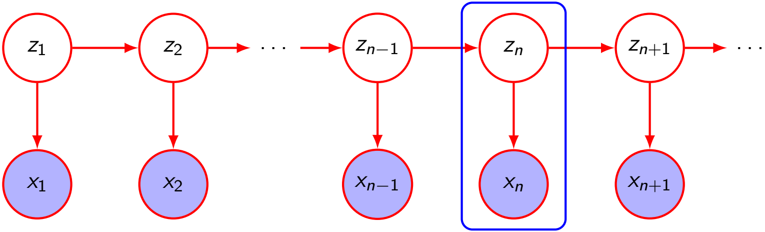


Solution: Viterbi algorithm

HMM Inference: Learning

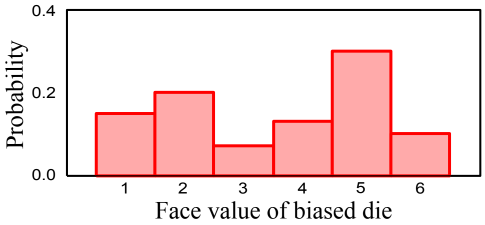
- estimate model parameters from data
- latent variables $\rightarrow$ Expectation Maximisation
- can be done efficiently with dynamic programming (Baum-Welch algorithm)

Emission Proabilities



- Emission: $p(x_n|z_n)$

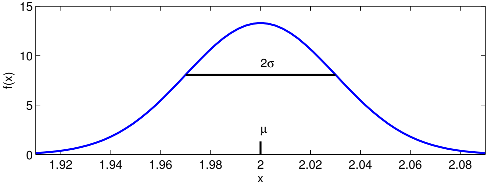
Discrete HMMs



Typical uses:

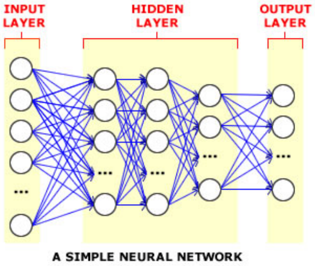
- DNA analysis
- Text analysis
- Speech Recognition (Vector Quantization)

Continuous HMMs



Typically Gaussian Distributions

Discriminative HMMs

x_n  s_n

- Artificial Neural Networks
- Support Vector Machines
- Deep Neural Networks

The output of the models are interpreted as probabilities

More about HMMs in:

DD2434	Machine Learning, Advanced Course (2nd period)
DD2380	Artificial Intelligence (1st period)
DT2119	Speech and Speaker Recognition (4th period)
EQ2340	Pattern Recognition (1st period)