Graphical Models for Machine Learning[1]

David Barber, Dmitry Adamskiy

University College London

[1] These slides accompany the book *Bayesian Reasoning and Machine Learning*. The book and demos can be downloaded from www.cs.ucl.ac.uk/staff/D.Barber/brml. Feedback and corrections are also available on the site. Feel free to adapt these slides for your own purposes, but please include a link the above website.

What are Graphical Models?

Probability and Models

- For many situations in science and machine learning we wish to make a model of the world.
- The world is described by a set of variables, for example ‘salary’, ‘house price’, ‘education level’, ‘marital status’
- Most generally, we need to deal with uncertainty and a model is then a distribution over these variables

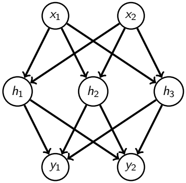
$$p(\text{'salary'}, \text{'house price'}, \text{'education level'}, \text{'marital status'})$$

Inference

- The model describes everything we know about the world and we can then query the model to answer questions of interest - this is called *inference*.

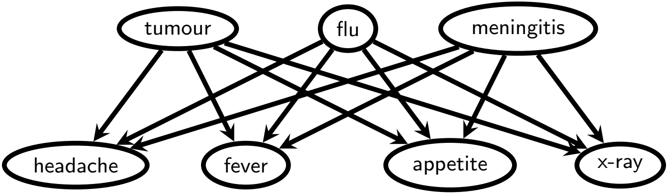
- For example, "Given that an individual has a high salary, is a graduate and is single, what is the distribution of house price".
- We can answer such inferential questions by using the rules of probability.

What are Graphical Models?



- Probability is a powerful framework. However, it is computationally expensive and we need to simplify things.
- Graphical Models are graph representations of *distributions*.
- Nodes in the graph represent variables and the edges represent statistical dependences between the variables, limiting which variables directly interact with each other.
- This greatly simplifies learning the parameters of these models and also in performing inference.
- WARNING! GMs are representations of distributions, whilst Neural Nets are graphical representations of *functions*. These are very different things.

Example: Medical Diagnosis



- Based on patient records, we can learn terms such as $p(\text{headache}|\text{tumour}, \text{flu}, \text{meningitis})$.
- We can use this model to infer for example what is the probability that a patient has meningitis given that they have a fever, headache, but no loss of appetite.
- These systems are used in practice with many thousands of variables.
- Have replaced rule-based expert systems (which are fragile and cannot deal with uncertainty).
- Tremendous potential to encode all known medical knowledge and help diagnoses.

Why Graphical Models?

- GMs are ways to deal with uncertainty and extremely useful way to communicate and describe probability distributions.
- GMs have been around a long time (they predate NNs)
- Developed in many different communities: Physics, Statistics, Computer Science, AI, Machine Learning, Engineering, Economics
- Applications in...
  - Computer Vision
  - NLP
  - Medical diagnosis
  - Genetics
  - Ranking systems
- ML students are expected to be familiar with the framework. Many models in ML are specified using GMs and the corresponding inference algorithms are widely used.

Some example applications

Burglar Problem

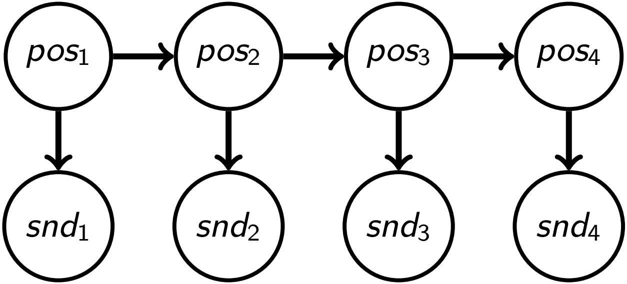
- You're asleep upstairs and are awoken by sounds from your kitchen downstairs.
- You know the layout of your kitchen and where the floorboard is likely to creak or where a burglar is likely to bump into an object in the dark.
- Based on listening to these bumps/creaks you try to infer where the burglar is.

Creaks and Bumps

Creak

Bump

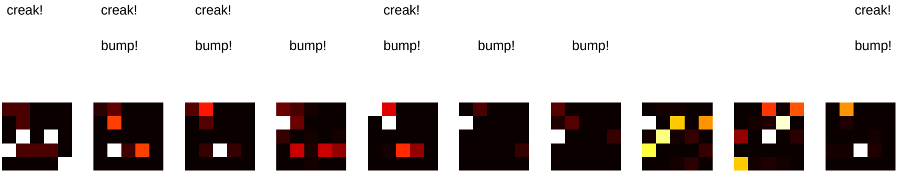
Burglar Model



pos - position in kitchen
snd – sound

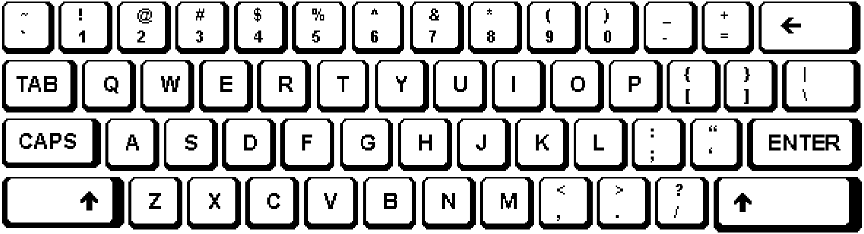
- The $p(snd|pos)$ describes the probability of hearing a certain sound, given the position (1 to 25) of the burglar.
- The term $p(pos_{t+1}|pos_t)$ describes how the burglar can move in the kitchen, for example moving only to a neighbouring square in a single timestep.
- This is an example of a Hidden Markov Model (HMM) and it is a simple Graphical Model (more on the details later).

Finding the Burglar



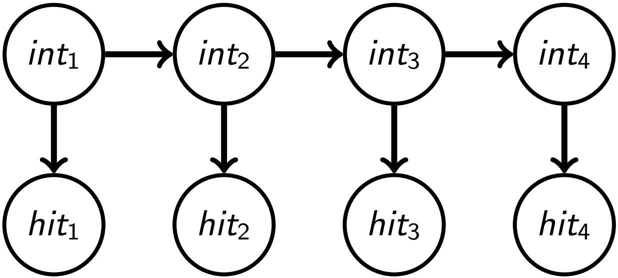
- Top Row : what we hear through time.
- Bottom Row : the distribution of where we think the burglar is based on all the information up to that timepoint.
- We are quite uncertain where the burglar is at the first time point since we don't have much information. As we gather more information, we start to get more confident as to where the burglar is – there are only a couple of likely positions where the burglar could be after 10 timesteps.

Stubby Fingers



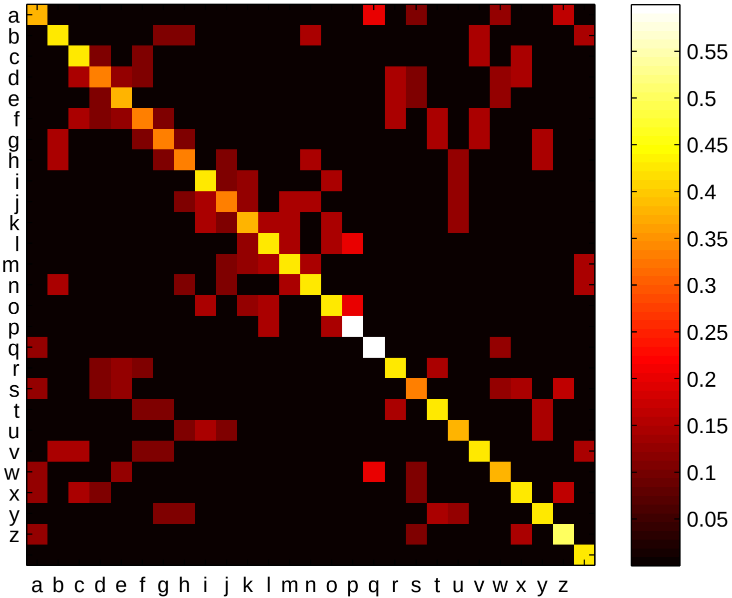
- Imagine that you have a standard keyboard, but that the typist has only a 70% chance of hitting the intended key, and a 30% chance of hitting a neighbouring key instead.
- This can be modelled by a term $p(hit|int)$
- We'll also assume that there is a distribution of next intended character, given current character $p(int_{t+1}|int_t)$ based on standard English.

Stubby Fingers

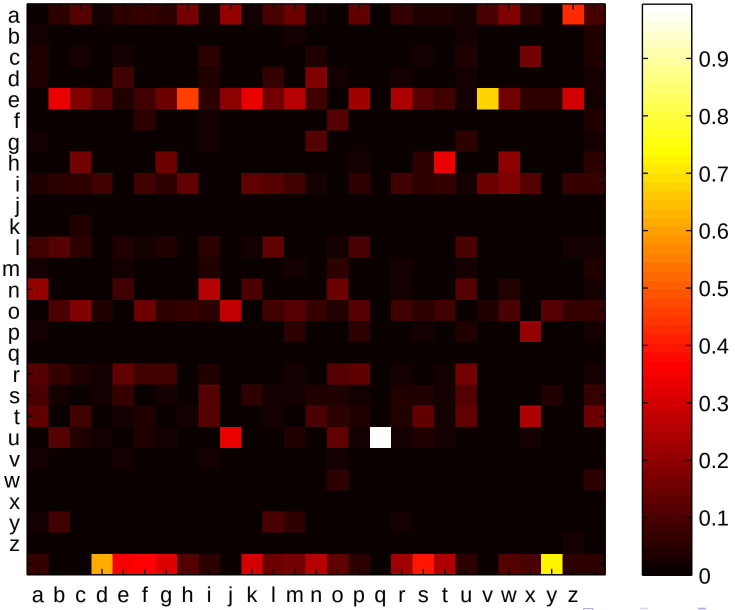


int - intended key
hit – hit key

Stubby Fingers: The term $p(int|hit)$



Stubby Fingers: language model $p(int_{t+1}|int_t)$

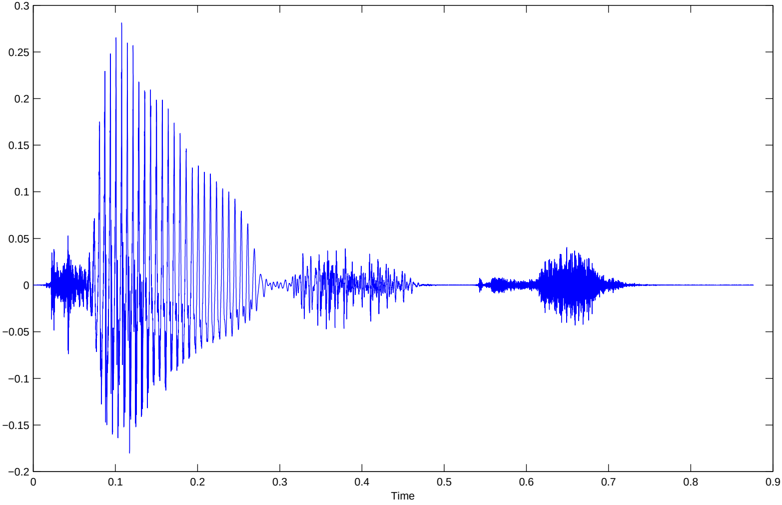


Stubby Fingers

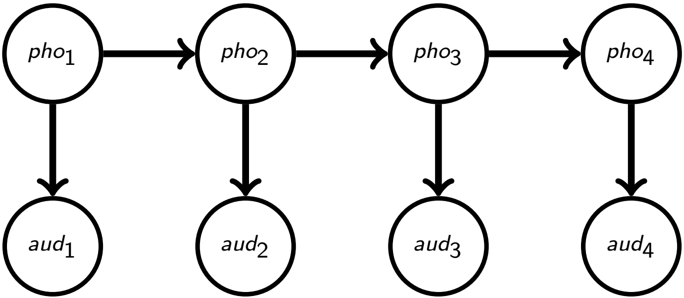
Given the typed sequence `kezrninh` what is the most likely word that this corresponds to?

- List the 200 most likely hidden sequences
- Discard those that are not in a standard English dictionary
- Take the most likely proper English word as the intended typed word
- Alternatively (gives the same result), for each word w in the dictionary, calculate $p(typed|w)$ and then choose the most likely w.

Speech Recognition: raw signal



Speech Recognition



- The term $p(aud|pho)$ describes how the audio signal would be observed if we knew which phoneme were generating the signal.
- The term $p(pho_{t+1}|pho_t)$ describes the distribution of phonemes at the next timestep, given we know the current phoneme – this information is provided by linguists.
- Given then the observed audio signal, we can infer the most likely sequence of phonemes that gave rise to the observed signal.

Summary

What are Graphical Models?

- Graphical Models are a marriage of graph and probability theory and used to both describe models and perform computations.
- They are ubiquitous across the sciences and used both in the physical sciences (modelling) and computer science (machine learning) and statistics.

What is this course about

- Learn how to specify graphical models.
- How to efficiently perform inference.
- How to deal with missing data or variables.
- How to perform approximate inference.
- Learn some specific models and their application in machine learning and other fields.

Free textbook and software

http://www.cs.ucl.ac.uk/staff/d.barber/brml/

- We will try to cover the material in the first 12 chapters
- Matlab and Julia software for Graphical Models and Machine Learning