6.867 Machine learning and neural networks

Tommi Jaakkola
MIT AI Lab

tommi@ai.mit.edu

Lecture 19: representation, graph models
Topics

• Representations of model structure
  – properties of representations
  – state diagrams vs. graph models
  – Bayesian networks
What is a good representation?

- Properties of good representations
  1. Explicit
  2. Compact
  3. Modular
  4. Permits efficient computation
  5. etc.
Representing the model structure

• Two possible representations of Markov models:

  1. in terms of state diagrams (nodes in the graph correspond to the possible values of the states)

     ![State Diagram](image)

     \[ s_0 \rightarrow s_1 \rightarrow s_2 \]

  2. in terms of variables (nodes in the graph are variables):

     \[ s_0 \rightarrow s_1 \rightarrow s_2 \]

• The representations differ in terms of what aspects of the model are made \textit{explicit}
Model structure cont’d

• Case 1: *sparse transition* structure

  1. State transition diagram is *explicit*

  ![State transition diagram](image)

  2. Representation in terms of variables leaves this *implicit*

  ![State transition diagram](image)
Model structure cont’d

• Case 2: successive states are *independent of each other*

  1. State transition diagram is fully connected

   ![State transition diagram](image)

  2. Representation in terms of variables is *explicit*

   \[ \begin{align*}
   &s_0 & s_1 & s_2 \\
   1 & & & \\
   2 & & & \\
   \end{align*} \]
Model structure cont’d

• Case 3: time series signals such as music may involve multiple relatively independent underlying processes operating at different time scales

1. State transition diagram (argh #$& …)

2. In terms of variables (graph model)

\[
\begin{align*}
& s_0^{(1)} \\
& \quad \downarrow \\
& s_0^{(2)} & \quad \rightarrow & s_1^{(2)} & \quad \rightarrow & s_2^{(2)} & \quad \rightarrow & s_3^{(2)} & \quad \rightarrow & s_4^{(2)} \\
& \quad \downarrow & \quad \rightarrow & \quad \rightarrow & \quad \rightarrow & \quad \rightarrow & \quad \rightarrow
\end{align*}
\]
Graphical models

- Graph representations of probability models in terms of variables are known as **graphical models**

A mixture model as a graphical model

- Different types of graph models differ in terms of how we represent dependencies and independencies among the variables
  1. Bayesian networks (natural for "causal" relations)
  2. Markov random fields (natural for physical or symmetric relations)
  3. etc.
Bayesian networks: examples

A Markov chain:

A hidden Markov model:
Qualitative inference

- The graph provides a qualitative description of the domain

\[ x_1 \] = first coin toss
\[ x_2 \] = second coin toss
\[ x_3 \] = same?
Qualitative inference

- The graph provides a qualitative description of the domain

\[ x_1 = \text{first coin toss} \]
\[ x_2 = \text{second coin toss} \]
\[ x_3 = \text{same?} \]

Marginal independence Induced dependence
Qualitative inference cont’d

• Just by looking at the graph, we can determine what we can and cannot ignore (why important?)

Marginal independence of “Earthquake” and “Burglary”

Diagram:

```
Earthquake  Burglary

Radio report  Alarm
```
Qualitative inference cont’d

• Induced dependence:

Earthquake  Burglary

Radio report  Alarm

• Explaining away:

Earthquake  Burglary

Radio report  Alarm
Two levels of description

- Graphical models need two levels of specification

1. Qualitative properties captured by a graph

   coin 1  
   \[ \xrightarrow{x_1 = \text{first coin toss}} \]  
   coin 2

   same or different

   \[ x_2 = \text{second coin toss} \]

   \[ x_3 = \text{same?} \]

2. Quantitative properties specified by the associated probability distribution

\[
P(x_1, x_2, x_3) = P(x_1) P(x_2) P(x_3|x_1, x_2)
\]

where, e.g.,

\[
P(x_1 = \text{heads}) = 0.5
\]

\[
P(x_3 = \text{same}|x_1 = \text{heads}, x_2 = \text{tails}) = 0
\]
More examples

- $i$ and $j$ correspond to the discrete choices in the mixture model
- $x$ is the (vector) variable whose density we wish to model
- We cannot tell what the component distributions $P(x|i)$ are based on the graph alone
In this case the choices of $i$ and $j$ and the output $y$ depend on the input $x$.

(The shaded variables denote observed values; we do not need to model the density over $x$)
More examples cont’d

In factorial HMMs, independent processes conspire to generate the observed output sequence.

In input-output HMMs, any observed sequence of outputs $y$ is accompanied by a corresponding sequence of inputs $x$.

– the model transforms any input sequence into an output sequence (markov?)
Graph model specification

- We need to address the following questions

  1. What is the graph semantics?
  2. What type of probability distribution can be associated with any specific graph?
  3. How can we exploit the graph in making quantitative inferences?
Graph semantics

- The graph captures *independence properties* among the variables.
- The independences can be read from the graph based on some notion of *graph separation*.

```
Course material  Lecturer  Test scores
Grade of student A  Course  Grade of student B
```

conditional independence
Graph semantics cont’d

• We have already seen the interesting cases...

Earthquake  Burglary
Alarm

marginal independence  induced dependence

• Note that the formal “graph separation” measure here must pay attention to the direction of the edges
Graph separation criterion (briefly)

- D-separation criterion (D for Directed edges):

  **Definition:** variables $x$ and $y$ are D-separated (conditionally independent) given $z$ if they are separated in the *moralized ancestral graph*

- Example: