1 Introduction

1.1 What is a probability model?

- Representation of a *joint distribution* over random configurations
  $\mathrm{P}(X_1 = x_1, ..., X_n = x_n)$

  That is, given a configuration $(x_1, x_2, ..., x_n)$,
  can determine its probability $\mathrm{P}(X_1 = x_1, ..., X_n = x_n)$

- Allows computation of

  - Marginal probabilities
    $\mathrm{P}(X_1 = x_1, ..., X_k = x_k)$;
    i.e. probability of a *partial* configuration
  - Conditional probabilities
    $\mathrm{P}(X_{k+1} = y_{k+1}, ..., X_n = y_n | X_1 = x_1, ..., X_k = x_k)$
  - Most likely completion
    $y^*_{k+1}, ..., y^*_n \;=\; \arg\max_{y_{k+1},\dots,y_n} \mathrm{P}(X_{k+1} = y_{k+1}, ..., X_n = y_n | X_1 = x_1, ..., X_k = x_k)$

1.2 Why are probability models important for AI?

Automating interpretation systems

- Computer perception (speech, vision)
- Natural language interpretation
- Expert systems
- Automated decision making
- Robotics

Plausible inference of hidden semantic structure from observable inputs

E.g.

input		hidden structure
word sequence	→	meaning
pixel matrix	→	object, relations
speech signal	→	phonemes, words
symptoms	→	illness
email message	→	spam/not spam

How to combine ambiguous, incomplete and conflicting evidence to draw reasonable conclusions?

Distinct from logical reasoning

- plausible inference:
  - non-monotonic: might change conclusions given more evidence
  - uncertain: conclusions are not guaranteed to be correct
    (but still want to do as well as possible)
- logical inference:
  - monotonic: once a conclusion is drawn it can never be retracted
  - certain: conclusions are certain given assumptions

1.3 How to build an interpretation system?

$$\text{observables} \rightarrow \boxed{\ ?\ } \rightarrow \text{hidden semantic structure}$$

Two key problems

1. need to represent facts about process that connects evidence to truth
2. need principles of evidence combination

In this course

We will represent uncertain knowledge using *probability theory*

Some alternatives we will not cover are

- fuzzy logic

- default logic
- rule-based systems
- Dempster-Shafer theory
- rough sets
- etc.

1.4 Building interpretation systems with probability models

Forward generative models

1. Represent knowledge with probability: forward generative models

$$
\begin{array}{c}
\downarrow \\
\boxed{\text{truth}} \\
\downarrow \\
\boxed{\text{evidence}}
\end{array}
\qquad\qquad
\begin{array}{c}
\text{world} \\
\downarrow \\
\text{sensor} \\
\downarrow \\
\text{observed measurement}
\end{array}
$$

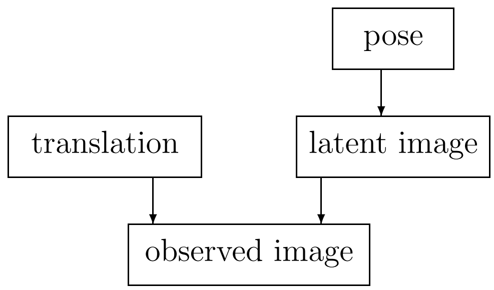
2. Principle of evidence combination: Bayesian inference

$$
\begin{aligned}
\text{conclusion} &= \arg \max_{\text{possible truth}} P(\text{possible truth}|\text{evidence}) \\
&= \arg \max_{\text{possible truth}} \frac{P(\text{evidence}|\text{possible truth})P(\text{possible truth})}{P(\text{evidence})} \\
&= \arg \max_{\text{possible truth}} P(\text{evidence}|\text{possible truth})P(\text{possible truth})
\end{aligned}
$$

1.5 Demos

Demo 1: Image normalization

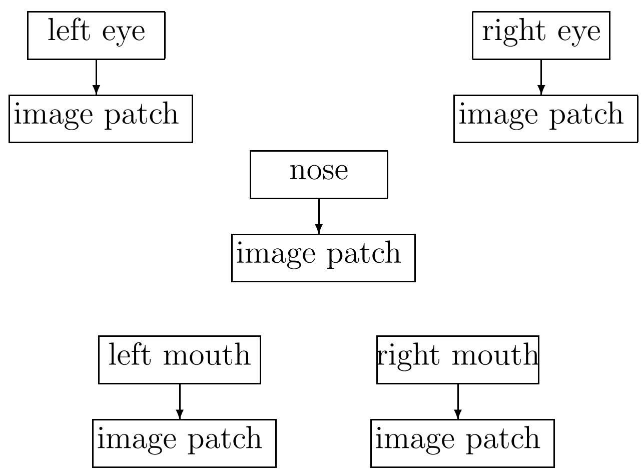
Bayesian inference



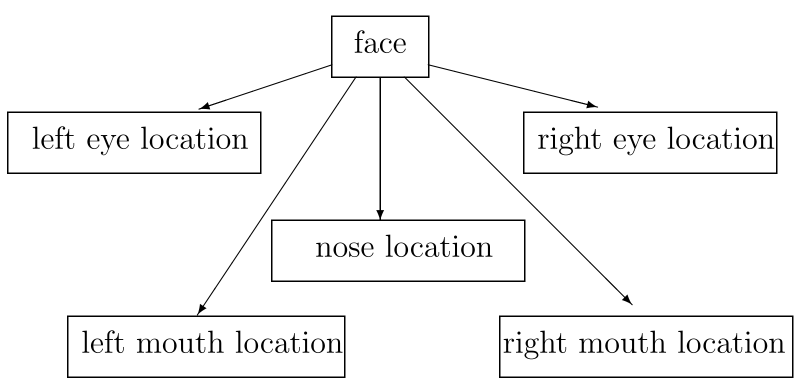
A time component is included to model image stabilization

Demo 2: Face tracking

Two models combined: low-level model



High-level model



Demo 3: Independent object tracking

Demo 4: Independent object tracking and object removal