# Homework 2

(Due on Friday before 09:50. Please submit your completed assignment to a TA by email or in-person.)

**AI policy.** The use of AI tools to complete the assignment is not permitted.

Please complete the questions below using the concepts and techniques introduced in the lectures.

## Markov-equivalence

**1.1.** Show that the graphs $A \leftarrow B \rightarrow C$ and $A \rightarrow B \rightarrow C$ are Markov-equivalent.

[*Hint*: think about the definition of Markov equivalence. Note also that the chain rule implies that $P(X|Y)P(Y) = P(Y|X)P(X)$ for all $X$ and $Y$.]

**1.2.** Show that the DAG $A \rightarrow B$ and the unconnected DAG $A \quad B$ are not Markov-equivalent.

[*Hint*: we simply need to find at least one joint distribution that can be factorized with respect to one DAG but not the other.]

## Categorizing blickets

A toy creature has three features: Color (either *red* or *blue*), Shape (either *round* or *square*), Texture (either *soft* or *hard*).

Some of these creatures emit a special kind of radiation that can be detected by a scientific instrument called a 'blicket detector'. For this reason these creatures are called 'blickets'.

You are trying to figure out what makes a creature a blicket. Armed with your blicket detector, you have tested whether four different individual creatures are blickets. Here are these 4 observations. Data $D = \{(o_i, y_i)\}_{i=1}^{4}$:

| Observation | Color | Shape | Texture | Blicket? |
|---|---|---|---|---|
| $o_1$ | red | round | soft | yes |
| $o_2$ | red | square | soft | yes |
| $o_3$ | blue | round | hard | no |
| $o_4$ | blue | square | soft | no |

On the basis of these data, one can evaluate hypotheses about what makes something a blicket. Suppose that hypotheses about "what makes something a blicket" are generated by this grammar:

$$H \rightarrow A \ (0.6) \qquad H \rightarrow (H \wedge H) \ (0.25) \qquad H \rightarrow (H \vee H) \ (0.15)$$

$$A \rightarrow \texttt{red} \mid \texttt{blue} \mid \texttt{round} \mid \texttt{square} \mid \texttt{soft} \mid \texttt{hard} \quad (\text{each } \tfrac{1}{6})$$

**2.1.** Compute the prior probability of the following hypotheses, and show every production rule used:

$$h_1 = \texttt{red}.$$
$$h_2 = (\texttt{red} \wedge \texttt{soft}).$$
$$h_3 = (\texttt{red} \vee \texttt{round}).$$
$$h_4 = ((\texttt{red} \wedge \texttt{soft}) \vee \texttt{round}).$$

**2.2.** Compare the prior probabilities $P(h_1)$, $P(h_2)$, $P(h_3)$, and $P(h_4)$. Explain why the probabilities differ.

**2.3.** Let us assume a likelihood function such that for each individual observation $o_i$, $P(o_i|H) = .1$ if the hypothesis makes an incorrect prediction about whether the creature is a blicket, and $P(o_i|H) = .9$ if the hypothesis makes a correct prediction. Then we have the likelihood:

$$P(D|H) = P(o_1|H)P(o_2|H)P(o_3|H)P(o_4|H).$$

Compute the likelihoods $P(D|h_1)$, $P(D|h_2)$, $P(D|h_3)$, and $P(D|h_4)$.

**2.4.** Compare the posterior probabilities $P(h_1|D), P(h_2|D)$, $P(h_3|D)$, and $P(h_4|D)$. Note that you don't need to compute the exact value of each probability, only their relative ranking. [*Hint*: The normalizing constant in Bayes' rule, $P(D)$, is the same for all $P(H_i|D)$.]

## Parsimony bias

Let $n_A(h)$ be the number of atomic predicates in a hypothesis $h$. Let $n_B(h)$ be the number of binary connectives, meaning uses of $\wedge$ or $\vee$.

For example,

$$h = (\texttt{red} \wedge \texttt{soft})$$

has

$$n_A(h) = 2, \qquad n_B(h) = 1.$$

**3.1.** For the grammar above, argue that every complete derivation satisfies

$$n_A(h) = n_B(h) + 1.$$

Hint: think of the derivation tree as a full binary tree.

**3.2.** Parsimony bias is the cognitive tendency to favor simpler explanations or models, also known as Occam's razor. Explain why PCFG creates a parsimony bias.

## (In)Tractability

The hypothesis space $\mathcal{H}$ contains all rules generated by the PCFG. Since the grammar is recursive, $\mathcal{H}$ is infinite.

**4.1.** Give three different valid hypotheses generated by the grammar that contain exactly one binary connective.

**4.2.** Give three different valid hypotheses generated by the grammar that contain exactly two binary connectives.

**4.3.** Roughly explain why the number of hypotheses grows quickly as the number of connectives increases.

**4.4.** The posterior normalizer is

$$P(D) = \sum_{h \in \mathcal{H}} P(D \mid h)P(h).$$

Explain why computing this sum exactly is difficult.

**4.5.** Even if many long hypotheses have small prior probability, why can we not always ignore all of them?

## Required Reflection

In 3–5 sentences, explain one mistake you initially made or one step in the homework that you found non-obvious. Your explanation should refer to a specific part of your solution.

**This reflection is required for getting any credit at all.**