4 Exercise Session 4

Exercises on probabilistic methods and reinforcement learning.

4.1 Naive Bayes

Classify `<single,light,one>` with Naive Bayes given the following examples:

```
Healthy: * <single,dark,one>
         * <single,light,two>
         * <double,light,one>

Virulent: * <single,dark,two>
          * <double,dark,one>
          * <double,light,two>
```

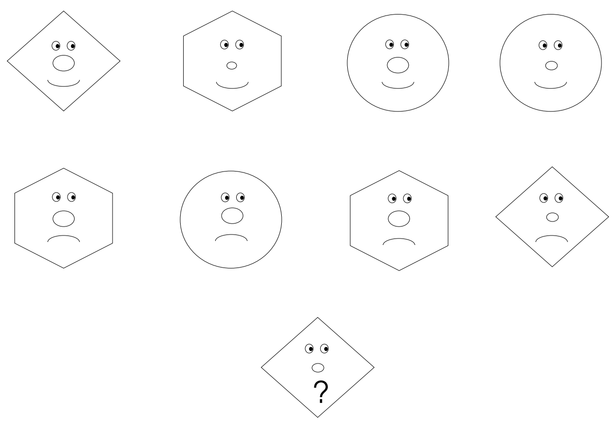
4.2 Text Classification

Classify the sentence “Cats eat mice and dogs bury bones.” given the following examples:

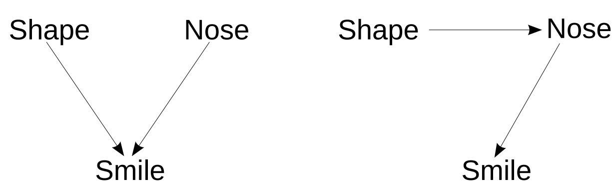
- Class A: “The cat crabs the curls off the stairs.”, “The cat curled under the stairs.”
- Class B: “It’s raining cats and dogs.”

and given Voc = {Cat, Crab, Curl, Stairs, Rain, Dog}.

4.3 Bayesian Networks



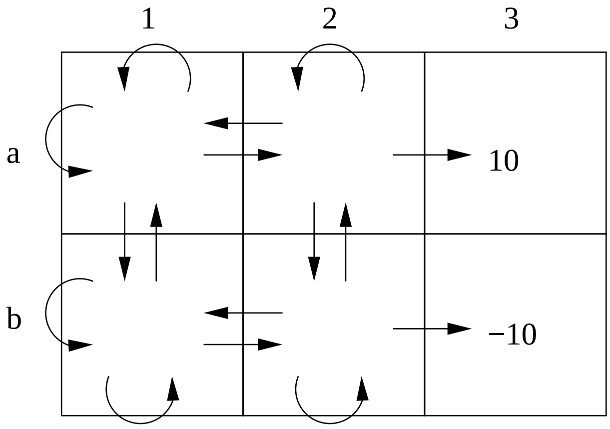
1. (Optional) Predict wether the last face is smiling or not using Naive Bayes.
2. Calculate all the probability tables for each of the Bayesian networks shown in the figure below according to the given examples-faces. Predict the class of the last face according to each of the networks.



3. Draw the Bayesian network that is equivalent to Naive Bayes.

4.4 Reinforcement Learning in a Nondeterministic Environment

Consider the following environment:



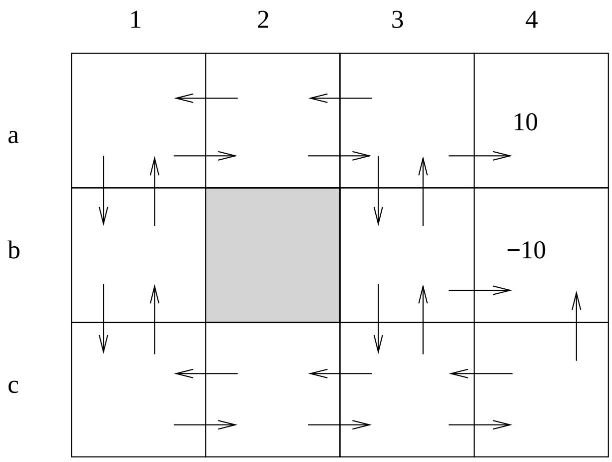
Entering the terminal states $a3$ and $b3$ gives a reward of 10 and a penalty of -10 respectively.

Write out the Bellman equations for the utility V^* of states for the optimal policy

1. when actions are deterministic
2. when actions are executed correctly in 70% of the cases, and a random other action is performed in other cases (bumping into wall results in staying in same state) using a discount factor $\gamma = 0.9$.

4.5 Q-Learning

Simulate Q-learning for a robot walking around in the following environment (b2 is a wall, entering b4 gives a penalty of -10, entering a4 gives a reward of 10).



Indicate Q-values after the following episodes, assuming the Q-values are initialized to 0 and using the "back-propagated" Q update rule (i.e. after getting in a goal state, updating the Q values in reverse order from goal to start, $\gamma = 0.9$).

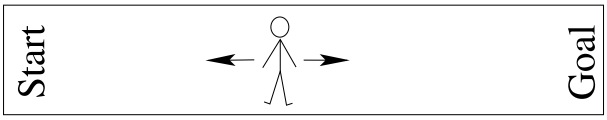
1) c2,c1,b1,a1,a2,a3,a4

2) a1,a2,a3,b3,b4

3) c4,c3,b3,a3,a4

Assume the robot will now use the policy of always performing the action having the greatest Q value. Indicate this policy on the drawing. Is it optimal?

4.6 Reinforcement Learning with Generalization



Consider the reinforcement learning problem shown above. The hallway is 10 meters long. The agent starts at the leftmost end of the hallway and receives a positive reward when it reaches the rightmost end of the hallway.

The position of the agent (i.e. its state) is described by a real number $\in [0:10]$ indicating its distance (in meters) from the leftmost end of the hallway (position 0). This kind of representation is known as a continuous state-space and it does not allow a straightforward use of a table-based representation of the Q-function.

Suppose the agent uses a step-size 1 and uses a discount factor $\gamma = 0.9$. Using TD(1), we have accurately estimated Q^*-values for 4 state-action pairs.

Example	State	Action	Q-Value
1	2	$\leftarrow$	4.30
2	9	$\rightarrow$	10
3	5	$\rightarrow$	6.56
4	5	$\leftarrow$	5.90

1. Use linear regression to generalize the estimated values to all possible positions in the hallway. (Hint: Use a separate function for each of the two available actions.)
2. Using these functions, compute (and interpret) the preferred action in both state "8" and state "1".

(Hint: For those who don't remember anything about high school math, you can also simply draw the functions.)

4.7 Using SAMIAM: Homework

SAMIAM is a tool to easily design and query Bayesian networks using a graphical interface. You can freely download it at http://reasoning.cs.ucla.edu/samiam/ (Samiam Release version). You can use it to gain more insight in the workings of Bayesian networks by reproducing within SAMIAM the networks of exercises 4.1 to 4.3.

1. Draw the networks of those three exercises, enter the probability tables, and see what results you obtain when making predictions.
2. Play around by modifying the probability tables (pay attention to the restrictions applying to probabilities and probability tables). How do the changes your make affect the predictions of your networks? Do all changes have the same impacts for the two networks of exercise 4.3.2?