

# Machine Learning in Complex Domains:

## Assignment 2

Daniel Deutsch, Dan Crankshaw, Ryan Cotterell

### 3 Problem Set

#### 3.1 Learning: Parameter Estimation

##### 3.1.5 Analytical Questions

1. Overfitting typically occurs in a complex model when the number of parameters is much larger than the number of observations. When you overfit the data, the model is more likely to exaggerate noise in the data instead of lessening its effect. In this problem, consider the case if the robot reaches location  $(i, j)$  at time  $t$ . If the robot does not observe a wall to the north when it should, the model will learn that it should never observe a wall to the north when at location  $(i, j)$  at time  $t$ . We can think of this mistake the model made as noise. The model learned the noise to be true, so whenever the robot is at  $(i, j)$  at time  $t$ , it will always predict the robot will observe no wall. This noise in the data has been exaggerated to always be true.

Therefore, when you are dealing with many parameters that are independent with respect to a specific variable, it is important to use parameter sharing so this situation does not happen and you lessen the effect of noise in the data. In this assignment, we say that a robot observing a wall is independent of the time it happens so that the model does not learn parameters for each time step and learn the noise into the model.

2. Suppose that in our model, the floor at location  $(i, j)$  is very sticky. When the robot tries to leave this location, it is more likely not to move compared to other locations on the map. When we use parameter sharing, we lose the information about moving in direction with respect to each position, and we can no longer encode that information. Parameter sharing makes the assumption that the probability of moving in a specific direction is the same for every position, and if that assumption is not true, parameter sharing cannot learn this model.
3. In order to combine these two approaches, you need to be able to detect an anomaly within the data, like a robot not being able to move as well in a specific location compared to the rest. A way to do this is to compute the statistics for every location individually. In our problem, this would be counting the number of times the robot is able to move in each direction in every location. Say that the sticky location is at  $(i, j)$ .

To detect that this is an anomaly, we will iterate over all of the positions, leaving 1 position out each time. Compute the average success of leaving the location and the rest of the locations. If the probability of leaving the sticky square is significantly different than from the rest of the locations, then leave that location out of the parameter sharing because it is an anomaly. Then you can learn parameters for each of the specific anomalies and shared parameters for the rest.

With this approach, you can learn the specific parameters for each space that you need to and general parameters for when it does not matter. In this way, you have taken the advantages of both solutions.

## **3.2 Inference**

### **3.2.1 Analytical Questions: Clique Tree**

- 1.
- 2.
- 3.

### **3.2.5 Empirical Questions: Message Passing**

- 1.
- 2.
- 3.
- 4.
- 5.
- 6.

### **3.2.6 Extra Credit: Max-Product Message Passing**

- 1.
- 2.

## **3.3 Bayesian Score for Bayesian Networks**

- 1.