David Hughes
davidralphhughes@college.harvard.edu
CS181-S18

Assignment #5
Due: 11:59pm April 18, 2018

Collaborators: Alexander Munoz

Homework 5: Graphical Models, MDPs

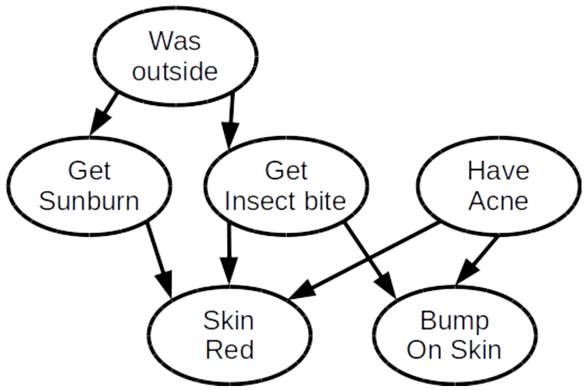
Introduction

There is a mathematical component and a programming component to this homework. Please submit your **tex, PDF, and Python files** to Canvas, and push all of your work to your GitHub repository. If a question requires you to make any plots, please include those in the writeup.

Bayesian Networks [7 pts]

Problem 1

In this problem we explore the conditional independence properties of a Bayesian Network. Consider the following Bayesian network representing a person's skin condition. Each random variable is binary (true/false).



The random variables are:

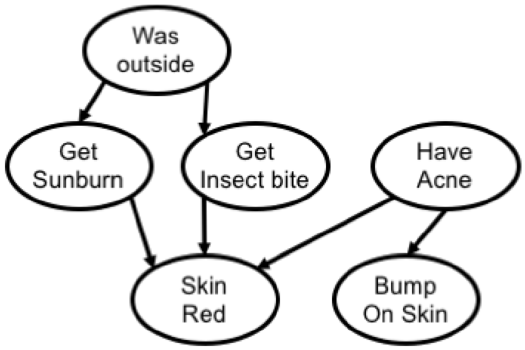
- Was Outside: Was the person outside?
- Get Sunburn: Did the person get sunburn?
- Get Insect Bite: Did the person get an insect bite?
- Have Acne: Does the person have acne?
- Skin Red: Is the skin red?
- Bump on Skin: Is there a bump?

For the following questions, $I(A, B)$ means that events A and B are independent and $I(A, B|C)$ means that events A and B are independent conditioned on C. Use the concept of d-separation to answer the questions and show your work.

1. Is $I(\textsf{Have Acne}, \textsf{Was Outside})$? If NO, give intuition for why.
2. Is $I(\textsf{Have Acne}, \textsf{Was Outside} \,|\, \textsf{Skin Red})$? If NO, give intuition for why.
3. Is $I(\textsf{Get Sunburn}, \textsf{Bump on Skin})$? If NO, give intuition for why.
4. Is $I(\textsf{Get Sunburn}, \textsf{Bump on Skin} \,|\, \textsf{Get Insect Bite})$? If NO, give intuition for why.
5. Suppose the person has taken a medicine to suppress their response to insect bites: they get red skin, but no bumps. Draw the modified network.
6. For this modified network, is $I(\textsf{Get Sunburn}, \textsf{Bump on Skin})$? If NO, give an intuition why. If YES, describe what observations (if any) would cause them to no longer be independent.

Solution

1. Yes they are independent.

2. No, they are not independent. If skin is red and have acne, there is a decreased probability of outside.
3. No, not independent. Having a sunburn increases the chances of being outside, which increases the chances of having an insect bite, and increases the probability of getting a bump on the skin.
4. Yes, independent.
5. 

6. Yes independent. If we observe skin red, they'd no longer be independent because having red skin and getting sunburned would decrease the probability of acne, and decrease the probability of a bump on the skin.

Kalman Filters [7 pts]

Problem 2

In this problem, you will implement a one-dimensional Kalman filter. Assume the following dynamical system model:

$$\begin{aligned} z_{t+1} &= z_t + \epsilon_t \\ x_t &= z_t + \gamma_t \end{aligned}$$

where z are the hidden variables and x are the observed measurements. The random variables ϵ and γ are drawn from the following Normal distributions:

$$\begin{aligned} \epsilon_t &\sim N(\mu_\epsilon, \sigma_\epsilon) \\ \gamma_t &\sim N(\mu_\gamma, \sigma_\gamma) \end{aligned}$$

where $\mu_\epsilon = 0$, $\sigma_\epsilon = 0.05$, $\mu_\gamma = 0$ and $\sigma_\gamma = 1.0$

You are provided with the observed data x and the hidden data z in kf-data.csv, and the prior on the first hidden state is $p(z_0) = \mathrm{N}(\mu_p,\sigma_p)$ where $\mu_p = 5$ and $\sigma_p = 1$

(a) The distribution $p(z_t|x_0...x_t)$ will be Gaussian $N(\mu_t, \sigma_t^2)$. Derive an iterative update for the mean μ_t and variance σ_t^2 given the mean and variance at the previous time step (μ_{t-1} and σ_{t-1}^2).

(b) Implement this update and apply it to the observed data above (do not use the hidden data to find these updates). Provide a plot of μ_t over time as well as a $2\sigma_t$ interval around μ_t (that is $\mu_t \pm 2\sigma_t$). Does the Kalman filter "catch up" to the true hidden object?

(c) Repeat the same process but change the observation at time $t = 10$ to $x_{t=10} = 10.2$ (an extreme outlier measurement). How does the Kalman filter respond to this outlier?

(d) Comment on the misspecification of dynamical system model for these data. Based on the previous two parts, how does this misspecification affect the predictions?

Solution

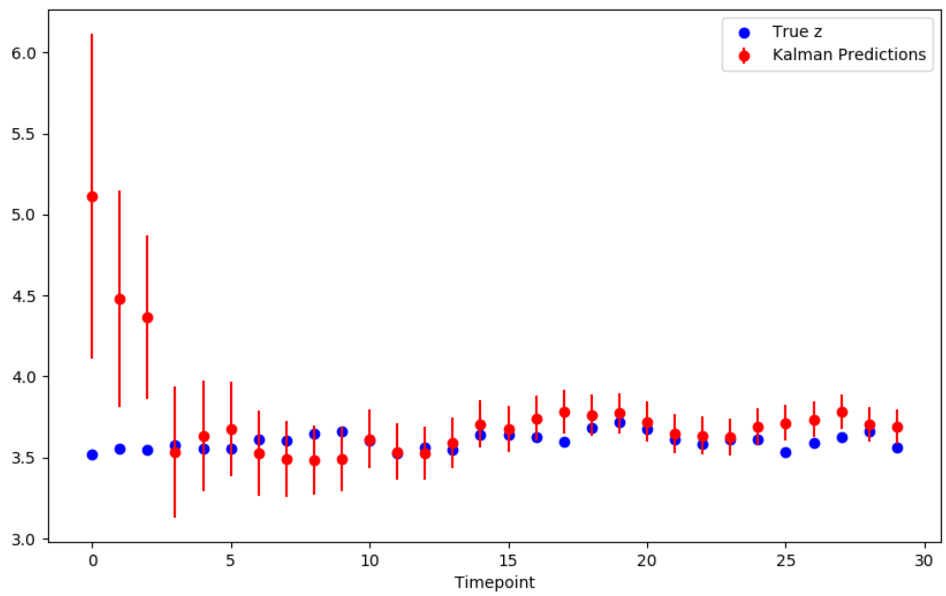
1. Complete the square to integrate:

$$\begin{aligned}
p(z_t|z_{t-1}) &\sim N(z_{t-1}, \sigma_\epsilon) \\
p(x_t|z_t) &\sim N(z_t, \sigma_\gamma) \\
p(z_t|x_{0:t}) &= \int_{z_{t-1}} p(z_t|z_{t-1,x_{0:t}}) dz_{t-1} \\
&= p(x_t|z_t) \int_{z_P t-1} p(z_t|z_{t-1}) p(z_{t-1}|x_{0:(t-1)}) dz_{t-1} \\
&\propto \int \exp\Bigg(-\frac{1}{2\sigma_\epsilon^2}(z_t - z_{t-1})^2 - \frac{1}{2\sigma_\gamma^2}(x_t - z_t)^2 - \frac{1}{2\sigma_{t-1}^2}(\overline{z}_{t-1} - z_{t-1}) dx_{t-1} \Bigg) \\
&= \exp\Bigg(-\frac{1}{2}\Bigg(z_t^2 \Bigg(\frac{1}{\sigma_\gamma^2} + \frac{1}{\sigma_\epsilon^2 + \sigma_{t-1}^2} \Bigg) - 2z_t \Bigg(\frac{\overline{x}_{t-1}}{\sigma_\epsilon^2 + \sigma_{t-1}^2} + \frac{x_t}{\sigma_\gamma^2} \Bigg) \Bigg) \Bigg)
\end{aligned}$$

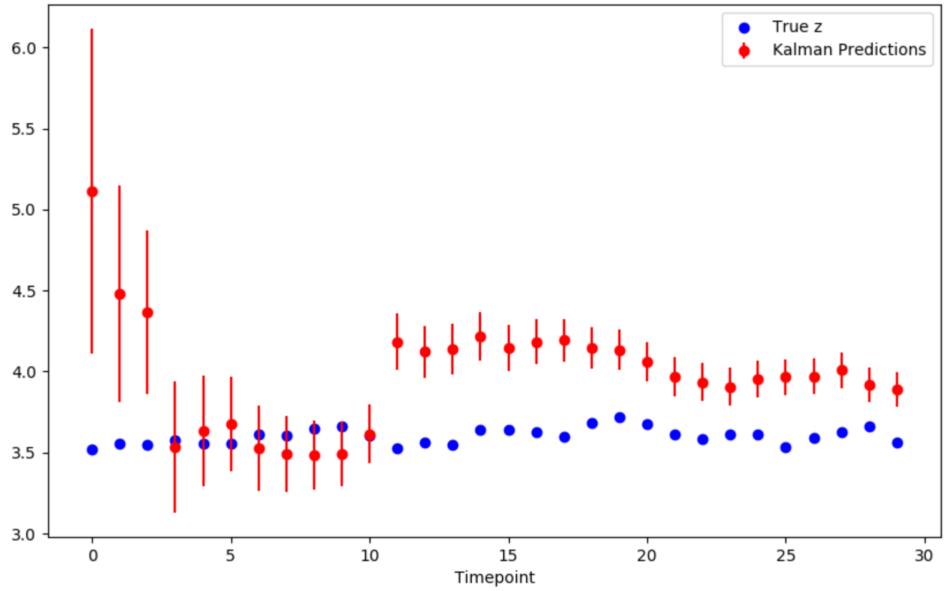
The updates are:

$$\overline{z}_t = \overline{z}_{t-1} + \frac{\sigma_\epsilon^2 + \sigma_{t-1}^2}{\sigma_\epsilon^2 + \sigma_{t-1}^2 + \sigma_\gamma^2}(x_t - \overline{z}_{t-1})$$
$$\sigma_t^2 = \sigma_\gamma^2 \frac{\sigma_\epsilon^2 + \sigma_{t-1}^2}{\sigma_\epsilon^2 + \sigma_{t-1}^2 + \sigma_\gamma^2}$$

2. The Kalman filter does catch up to the true hidden object.



3. The Kalman filter is drastically affected by the single outlier.



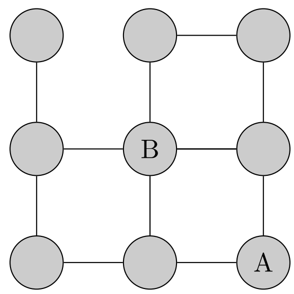
4. The model is mis-specified because the single outlier drastically affects the Kalman filter predictions, however the error bars do not account for this change. Looking at the formulae for the updates in part 2, we can see the fraction determines the power of the update: the larger this fraction, the more we will update based on new observations. We tune increasing σ_γ to rely less on the new update and decrease sensitivity to outliers.

Markov Decision Processes [7 pts]

Problem 3

In this problem we will explore the calculation of the *MDP value function* V in a 2D exploration setting, without time discounting and for a finite time horizon.

Consider a robot navigating the following grid:



The robot moves in exactly one direction each turn (and must move). The robot's goal is to maximize its score in the game after T steps. The score is defined in the following way:

- If the robot's movement attempts to move it off the grid, then the robot loses a point (-1) and stays where it is.
- If the robot ends its motion on node A, it receives 10 points, and if it moves onto node B, it receives -100 points. Otherwise, it receives 0 points.

1. Model this as a Markov decision process: define the states S, actions A, reward function $r : S \times A \mapsto \mathbb{R}$, and transition model $p(s' \mid s, a)$ for $s', s \in S$ and $a \in A$. For now, assume that the robot's actions execute perfectly: if the robot tries to move in a particular direction, it always succeeds in doing so.
2. Consider a *random policy* π, where in each state the robot moves uniformly at randomly in any of its available directions (including off the board). For every position on the grid calculate the value function, $V_t^\pi : S \mapsto \mathbb{R}$, under this policy, for $t = 2, 1$ steps left to go. You can find LaTeX code for the tables in the solution template. Note that you should have 2 tables, one for each time horizon.
3. Now assume that the robot plays an *optimal policy* π_t^* (for t time steps to go). Find the optimal policy in the case of a finite time horizon of $t = 1, 2$ and give the corresponding MDP value functions $V_t^* : S \mapsto \mathbb{R}$, under this optimal policy. You can indicate the optimal policy for each time horizon on the corresponding V_t^* table via arrows or words in the direction that the robot should move from that state.
4. Now consider the situation where the robot does not have complete control over its movement. In particular, when it chooses a direction, there is a 80% chance that it will go in that direction, and a 10% chance it will go in the two adjacent (90° left or 90° right) directions. Explain how this changes the elements S, A, r, and $p(s' \mid s, a)$ of the MDP model. Assume the robot uses the same policy π_t^* from the previous question (now possibly non-optimal), and write this as π_t, and tie-break in favor of N, then E, then S then W. Give the corresponding MDP value functions $V_t^\pi : S \mapsto \mathbb{R}$, for this policy in this partial control world, for $t = 2, 1$ steps left to go. Is the policy still optimal?

Solution

1. The states S are the 9 different nodes. The actions A are up, down, left, and right. The reward function is -1 if the action ends in a state off the map, +10 if the action ends in state A, -100 if the action ends in state B, and 0 otherwise. If the action will take you off the map, the transition probability is 0, and 1 otherwise.

2. To calculate with one timepoint to go, we average over the probabilities:

$$V^{\pi}_{t=1}(s) = r(s, \pi(s))$$

-0.75	-25.5	-0.5
-0.25	0	-22.75
-0.5	-22.75	-0.5

With two timepoints left, we add the expected current reward with the expected future reward:

$$V^{\pi}_{t=2}(s) = r(s, \pi(s)) + \sum_{s' \in S} p(s'|s, \pi(s)) V^{\pi}_{t-1}(s')$$

-1.375	-38.375	-12.81
-0.575	-17.81	-28.69
-0.65	-28.69	-12.13

3.

t=1:

$\downarrow (0)$	$\rightarrow (0)$	$\downarrow (0)$
$\uparrow (0)$	$\uparrow (0)$	$\downarrow (+10)$
$\uparrow (0)$	$\rightarrow (+10)$	$\uparrow (0)$

t=2:

$\downarrow (0)$	$\rightarrow (0)$	$\downarrow (+10)$
$\uparrow (0)$	$\rightarrow (+10)$	$\downarrow (+10)$
$\rightarrow (+10)$	$\rightarrow (+10)$	$\uparrow (+10)$

4. States S, actions A, and reward function r would be identical as in previous parts. The difference now is in the transition probabilities. Specifically, while trying to go from state s to state s', there is an 80% chance of transitioning to s', a 10% chance of transitioning 90 degrees left, and 10% chance of transitioning 90 degrees right.

t=1:

$\downarrow (-0.2)$	$\rightarrow (-10.1)$	$\downarrow (-0.1)$
$\uparrow (-10.1)$	$\uparrow (0)$	$\downarrow (-9.1)$
$\uparrow (-0.1)$	$\rightarrow (-9.1)$	$\uparrow (-0.1)$

t=2:

$\downarrow (-9.31)$	$\rightarrow (-10.41)$	$\downarrow (-8.4)$
$\uparrow (-11.27)$	$\rightarrow (-9.2)$	$\downarrow (-10.09)$
$\rightarrow (-8.4)$	$\rightarrow (-10.09)$	$\uparrow (-8.3)$

Problem 4 (Calibration, 1pt)

Approximately how long did this homework take you to complete? 10 hours

- Name: David Hughes
- Email: davidralphhughes@college.harvard.edu
- Collaborators: Alexander Munoz