UNIVERSITY OF SASKATCHEWAN
Department of Computer Science
176 Thorvaldson Building
110 Science Place, Saskatoon, SK, S7N 5C9, Canada
Telephine: (306) 966-4886, Facimile: (306) 966-4884

Assignment 5

Bayesian Netowrks, and Linear Learning Models

Date Due: 7 December 2018, 11:59pm **Total Marks: 33**

General Instructions

- **This assignment is individual work.** You may discuss questions and problems with anyone, but the work you hand in for this assignment must be your own work.
- Each question indicates what to hand in.
- Do not submit folders, or zip files, even if you think it will help.
- **Assignments must be submitted to Moodle.**

Version History

- **03/12/2018**: released to students

Overview

In this assignment, we'll exercise a few ideas, but there is no significant programming. This assignment should take you about 2-4 hours.

Questions 1,2 can be done on paper (then scanned and submitted electronically), or using a word processor. Whatever you decide, make it legible, keep it simple, and prefer PDF and JPG file formats.

Questions 3 and 4 are very short exercises exploring linear classifiers. Most of the work is done for you, but it has to be completed using Jupyter Notebook. This is a way of working with scripts in various languages, combining documentation, code, and output.

A version of Jupyter Notebook is available on Linux machines in the lab. Furthermore, the department has a cloud service set up for Jupyter Lab, which you can use:

- `https://trux.usask.ca`

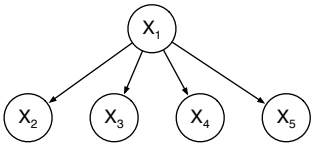
The only trick will be getting the files you need on the lab filesystem.

Jupyter Notebook is easy to install on your own computers, especially if you use the Anaconda 3 installation tools for Python 3. If you installed Python using Anaconda 3, it may already be installed for you. If not, Google for help. This might take an extra 15 or 20 minutes.

Question 1 (12 points):

Purpose: To reinforce the concepts of Bayesian networks.

Consider the Bayesian network given below.



(a) (5 points) What are the Conditional Probability Distributions implied by the network diagram above? List them using the notation $P(\ldots)$. You do not need to indicate any probabilities. Just provide the notation.

(b) (2 points) Assume that each variable X_i has 10 domain values. How many entries in each Conditional Probability table that you listed? In other words, how many numbers would be required if you were to fill in each table (which you thankfully don't have to do). What's the total number of entries, when you add up all the entries for all the CPDs?

(c) (1 point) Express the Joint Probability Distribution in terms of the Conditional Probability Distributions you outlined above.

(d) (4 points) Derive a formula for the query $P(X_1|X_2, X_3, X_4)$.

What to Hand In

Hand in a document containing the answers to the above questions. You may submit a text document, or you can complete the work on paper, and submit an image of it (scanned, or captured by your phone). Before you submit, make sure it's legible.

- The answers to the questions above.

Be sure to include your name, NSID, student number, and course number at the top of all documents.

Evaluation

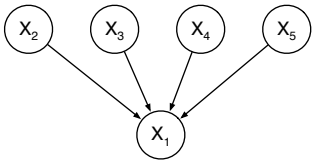
- 4 marks: You described the Conditional Probability tables using the $P()$ notation, indicating the correct dependencies as implied in the network structure.
- 2 marks: You calculated the correct number of entries for each CPD, and in total.
- 1 mark: Your JPD was correctly defined in terms of the CPDs.
- 4 marks: Your formula was derived correctly.

If the submission is not legible, the marker may deduct up to 100% of the grade.

Question 2 (12 points):

Purpose: To reinforce the concepts of Bayesian networks.

Consider the Bayesian network given below.



(a) (5 points) What are the Conditional Probability Distributions implied by the network diagram above? List them using the notation $P(\ldots)$. You do not need to indicate any probabilities. Just provide the notation.

(b) (2 points) Assume that each variable X_i has 10 domain values. How many entries in each Conditional Probability table that you listed? In other words, how many numbers would be required if you were to fill in each table (which you thankfully don't have to do). What's the total number of entries, when you add up all the entries for all the CPDs?

(c) (1 point) Express the Joint Probability Distribution in terms of the Conditional Probability Distributions you outlined above.

(d) (4 points) Derive a formula for the query $P(X_1|X_2, X_3, X_4)$.

What to Hand In

Hand in a document containing the answers to the above questions. You may submit a text document, or you can complete the work on paper, and submit an image of it (scanned, or captured by your phone). Before you submit, make sure it's legible.

- The answers to the questions above.

Be sure to include your name, NSID, student number, and course number at the top of all documents.

Evaluation

- 4 marks: You described the Conditional Probability tables using the $P()$ notation, indicating the correct dependencies as implied in the network structure.
- 2 marks: You calculated the correct number of entries for each CPD, and in total.
- 1 mark: Your JPD was correctly defined in terms of the CPDs.
- 4 marks: Your formula was derived correctly.

If the submission is not legible, the marker may deduct up to 100% of the grade.

Question 3 (5 points):

Purpose: To work experiment with a simple implementation of a Linear Classifier.

On the Moodle Assignment page, you'll find the files `linclass.py` and `A5Q3.ipynb`. Download them both.
Using Jupyter Notebook, open `A5Q3.ipynb`. Read through the document.
Complete the TO DO items in the notebook:

(a) (2 points) TO DO: Choose a different line
(b) (1 point) TO DO: Add a point to the line
(c) (2 points) TO DO: Increase the number of learning steps

What to Hand In

The Jupyter Notebook `A5Q3.ipynb` with the TO DO items complete.
Be sure to include your name, NSID, student number, and course number at the top of all documents.

Evaluation

1. (2 marks) Your new line separates the original data.
2. (1 mark) The point you added was misclassified by the line from the previous task.
3. (2 marks) You increase the number of learning steps, and the fitted line separates the two classes.

Question 4 (4 points):

Purpose: To work experiment with a simple implementation of a Linear Classifier.

On the Moodle Assignment page, you'll find the files `linclass.py` and `A5Q4.ipynb`. Download them both.

Using Jupyter Notebook, open `A5Q4.ipynb`. Read through the document.

Complete the TO DO items in the notebook:

(a) (2 points) TO DO: Increase the number of learning steps (linear classifier)

(b) (2 points) TO DO: Increase the number of learning steps (logistic classifier)

What to Hand In

The Jupyter Notebook `A5Q4.ipynb` with the TO DO items complete.

Be sure to include your name, NSID, student number, and course number at the top of all documents.

Evaluation

(a) (2 marks) You increased the number of learning steps, and changed the `alpha` parameter. Your linear classifying line does not perfectly classify the data.

(b) (2 marks) You increased the number of learning steps, and changed the `alpha` parameter. The Logistic Classifier gets all but one data point correctly classified.