Assignment 3: Graphical Models

Due March 23 at 11:59pm
70 marks total

This assignment is to be done individually.

Important Note: The university policy on academic dishonesty (cheating) will be taken very seriously in this course. You may not provide or use any solution, in whole or in part, to or by another student.

You are encouraged to discuss the concepts involved in the questions with other students. If you are in doubt as to what constitutes acceptable discussion, please ask! Further, please take advantage of office hours offered by the instructor and the TA if you are having difficulties with this assignment.

DO NOT:

- Give/receive code or proofs to/from other students
- Use Google to find solutions for assignment

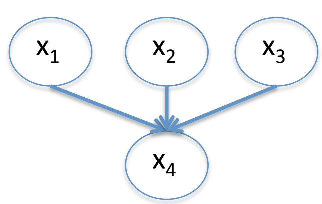
DO:

- Meet with other students to discuss assignment (it is best not to take any notes during such meetings, and to re-work assignment on your own)
- Use online resources (e.g. Wikipedia) to understand the concepts needed to solve the assignment

---

Question 1 (10 marks)

Consider learning the parameters for the Bayesian network shown in the figure.



Suppose we have a training set $(\boldsymbol{x}^1, \boldsymbol{x}^2, \ldots, \boldsymbol{x}^N)$, where each $\boldsymbol{x}^i = (x_1^i, \ldots, x_4^i)$ is a vector containing values for all 4 random variables in the network.

1. Write down the likelihood and the log-likelihood of the training data given a parameter setting of the Bayes net. Please use the following notation.

(a) θ_{ijk} = the conditional probability of node i taking on value k given that the parents of i are in state j. In our example, $i = 1, \ldots, 4$, and $k = 1, \ldots, L$. If $i = 1, 2, 3$, then $j = 0$, that is j is just a dummy index since the first three nodes have no parents. If $i = 4$, then $j = 1, \ldots, L^3$.

(b) n_{ijk} = the number of training cases where node i takes on value k and the parents of i are in state j.

2. Show that with binary nodes ($L = 2$) the maximum likelihood parameter setting are the conditional frequencies observed in the data:

$$\theta_{ijk} = \frac{n_{ijk}}{\sum_{k'} n_{ijk'}}$$

3. Show that with any number of states L, the maximum likelihood parameter setting are the conditional frequencies observed in the data. (Hint: use Lagrange multipliers).

The maximum likelihood result holds generally for a Bayes net with discrete values, I have given you a specific example only for the sake of concreteness.

Question 2 (10 marks)

Question 8.16 in PRML.

Question 3 (50 marks)

One use of generative models is to perform feature selection for classification or regression. In this exercise you will experiment with Bayes net learners for this purpose. The idea is to redo assignment 1 on linear regression, but first using a Bayes net to find relevant variables. The instructions are pretty long, but that's because we use other people's programming, so the actual time for the assignment shouldn't be longer than that for the other assignments.

Prepare the data (or use the already prepared data)

The dataset is again the `AutoMPG` dataset from the UCI repository. We use the normalized features as in Assignment 1. We include 4 data files for reading into Tetrad. These are basically 4 of the design matrices you used in Assignment 1, without the bias columns and with the target labels. If you want to build the data files yourself, check the description below and the sample code assg3.m. For training, we use the first 100 points, and for testing the remainder. Here are the details of how we built the data files.

1. Start with the design matrix with polynomial basis functions for each of degree 1,3 so we have a total of 2 design matrices.

2. Export the design matrix as a Tetrad readable data matrix, where the columns are labelled with the basis functions (features) as well as the target label (fuel effiency) and the rows correspond to data points. Omit the constant 1 dummy column for the bias weight.
3. Repeat the procedure of Assignment 1, Gaussian basis functions, part 1, to build 2 design matrices with 5,15 basis functions respectively. Export these design matrices to Tetrad format.

Learn Bayes Nets and Find the Markov Blanket

1. Go to Tetrad `http://www.phil.cmu.edu/projects/tetrad_download/`. For info about the Tetrad program, download the manuals (the old and the new, the new seems to be missing the beginning). Unless I specify otherwise, you should use the default settings. If these don't work for you, please check the documentation first, then e-mail Majid, or ask me in class.
2. Launch Tetrad 4.3.9-14. This requires an updated version of JRE. If you have trouble launching this from the browser, try downloading a jar file and starting that. (Warning: the latest version Tetrad 4.3.10 seems to be buggy, don't use that.)
3. For each of the 4 design matrix data sets, do the following.
   (a) Go to the "Template" Item on the main menu and select "Search from Loaded Data".
   (b) Click on the Data1 box, select Data Wrapper, select Load data from File. Load your design matrix.
   (c) Click on the Search1 box. Choose GES search to learn a Bayes net for the data table.
   (d) Find the Markov blanket for the target variable in the GES search result. If you are unsure how to read the undirected edges, select "DAGS from pattern" and use any of the DAGs that Tetrad offers you. (Note: All DAGs should have the same Markov blanket).
4. Write down the Markov blanket for each dataset. I suggest using a table where the rows are the type of basis functions used, the columns are the maximum degree/number of basis functions, and each cell contains a list of variables, which are those in the Markov blanket.

   Do you have any observations about the Markov blankets? For instance, do they grow or shrink with the number of basis functions?

Use the Markov Blankets

For each of the 4 datasets, do the following. (Basically, repeat Assignment 1.)

1. Discard variables that are *not* in the Markov blanket to form a new smaller dataset. (Tetrad may have a utility for this. Otherwise it can easily be done in Excel or such.)

2. Apply L2-regularized regression using the smaller Markov blanket data using $\lambda = \{0, 0.01, 0.1, 1, 10, 100, 1000\}$. Use cross-validation to decide on the best value λ^*, apply regression to find the best set of weights for the best λ^*, and compute the RMS error on the training data for this set of weights and λ^*. (Variables not in the Markov blanket effectively have 0 weight and can be excluded.)

Finally, compare linear regression with and without prior feature selection.

1. **Produce a plot or table showing the test set error of your newly learned regression models for the 4 test data sets. Include a comparison with the linear regression test error without Bayes net feature selection.** Does feature selection change the performance on the test data set? For better or for worse?

2. Include any other informative comparison observations. For instances, does the Bayes net learner include in the Markov blanket all the variables that are assigned high weights in the linear regression models you learned in Assignment 1? Does it include only such variables in the Markov blanket? You may describe other observations, but please be brief, about half a page or so.

Submitting Your Assignment

You should create a report with the answers to questions and figures described above in PDF format. Make sure it is clear what is shown in each figure. **DO NOT INCLUDE SOURCE CODE.**

Submit your assignment using the online assignment submission server at: `https://submit.cs.sfu.ca/`