CS 5824/ECE 5424: Advanced Machine Learning
Assignment 2

Fall 2022

Due Date: October 25, 2022, 11:59pm EST

Introduction

This homework will cover decision trees, support vector machines (SVMs), and probabilistic graphical models for ML. These exercises will help you better understand and intuit more advanced ideas in ML so as to potentially employ them in future ML endeavors.

The homework consists of two sections, Section A and Section B. Only Section A questions are given in this document. Section B is a Python Notebook that you will have received along with this document. You are expected to solve both the questions in the notebook and in this problem set, i.e., complete all questions from both sections A and B.

- The homework is due at 11:59 PM on the due date. We will be using Canvas for collecting homework assignments. For section A, feel free to submit a PDF, either of a scanned copy of your handwritten solution or a typed solution (using Microsoft Word, Latex, etc). Contact the TAs if you have technical difficulties in submitting the assignment. Section A homework should be submitted as a single pdf using the name convention **yourFirstName-yourLastName.pdf**.
- We encourage you to not submit late so that you don't accrue late days. Please declare the number of late days used for this homework at the top of your report. **Refer to the late days policy on the canvas page for more information.** *TLDR: HWs are to be done individually, and each student gets up to 5 late days to use as per their discretion. No more than up to 3 late days can be allocated to a specific HW.*
- For each question, please include all necessary calculation steps. If the result is not an integer, round your result to 3 decimal places.
- Please use the discussion section on Piazza (`https://piazza.com/class/l74tedj5el86tp`) to ask questions about the homework. Also, feel free to e-mail us at `cs5824-g@cs.vt.edu` and come to office hours.

Section A [60 pts]

You are expected to submit your solutions to this section as a PDF file. LaTeX is strongly preferred, but is not mandatory. The solutions need to be neat and legible. You should provide reasoning for your solution and show all your work. Except for obvious solutions, points will be deducted if there is no reasoning provided.

Problem 1. Decision Trees (20 points total)

It's Saturday afternoon. You are kind of bored but not bored enough to know whether you should absolutely go out to eat or not. Luckily, you keep a handy journal of all the decisions you have made on previous Saturdays based on four different factors: how much homework you have, how busy traffic is, how hungry you are, and whether your best friend Lauren Ipsum is available. You want to construct a decision tree from this dataset to help you make a decision today. Your dataset is as follows:

Homework	**Traffic**	**Hunger**	**Lauren**	**Go out?**
Much	Busy	A little	Available	No
Much	Busy	A little	Not available	No
Normal	Busy	A little	Available	Yes
None	OK	A little	Available	Yes
None	Chill	A lot	Available	Yes
None	Chill	A lot	Not available	No
Normal	Chill	A lot	Not available	Yes
Much	OK	A little	Available	No
Much	Chill	A lot	Available	Yes
None	OK	A lot	Available	Yes
Much	OK	A lot	Not available	Yes
Normal	OK	A little	Not available	Yes
Normal	Busy	A lot	Available	Yes
None	OK	A little	Not available	No

Table 1: "What did I do past Saturdays?"

Q1. (5 points) You vaguely remember that using the information gain ratio can help remedy one of information gain's downsides when constructing a decision tree. Will doing so result in a significant change with regards to this dataset? Why or why not?

Q2. (5 points) Compute the gain ratio from each feature for the first split of your decision tree using entropy as your purity criterion (C4.5 algorithm). Which feature should be your first split?

Q3. (8 points) Complete your decision tree. Include the tree and corresponding computations in your write-up; your tree could be either a digitally rendered visualization or a photo of a manual graph.

Q4. (2 points) Today, you have a normal amount of homework and are very hungry. However, traffic is busy because it's Hokie game day, and Lauren is not available to hang out with you. According to your decision tree, are you going out?

Problem 2. Support Vector Machines (15 points total)

Given a linearly separable dataset where each sample is either from class $y_i = 1$ or class $y_i = -1$, a linear SVM can always find the separating hyperplane with the maximum margin. Assuming we know the two nearest training samples from different classes are $\mathbf{x}_j, y_j = 1$ and $\mathbf{x}_k, y_k = -1$ and the parameters for the linear SVM are $\mathbf{w}$ and b, the linear SVM tries to maximize the margin as follows:

$$\arg\max_{\mathbf{w},b} d = \arg\max_{\mathbf{w},b} \left(d^+ - d^-\right) = \arg\max_{\mathbf{w},b} \left(\frac{\mathbf{w}^T\mathbf{x}_j}{\|\mathbf{w}\|_2} - \frac{\mathbf{w}^T\mathbf{x}_k}{\|\mathbf{w}\|_2} \right)$$

where d^+ (d^-) denotes the distance between $\mathbf{x}_j$ ($\mathbf{x}_k$) and the separating hyperplane.

Q1. (5 points) The above optimization problem is often too hard to solve. Derive an alternative optimization problem (including an objective function and constraints) of the "hard-margin" linear SVM (*i.e.*, SVM without slack variables). Justify your answer.

Q2. (5 points) In practice, it is often the case that data points cannot be well-separated via a hard margin. Can you provide a solution to solve this problem? What does the new optimization problem look like?

Q3. (5 points) Is it possible to apply SVM to multi-class classification? Justify your answer.

Problem 3. Probabilistic Graphical Models (25 points total)

Q1. (10 points) Given N random variables $x_1, x_2, ..., x_N$, you want to graph a Bayesian network showing the joint distribution of all N variables. Assume no independence relations.

1. (4 points) How many distinct networks could you possibly graph in terms of N? Justify your answer.
2. (6 points) The N random variables each have their own number of states s ($s \in \mathbb{Z}$, $s \geq 2$), or the number of discrete values they can take on. The variables $x_1, x_2, ..., x_N$ have $s_1, s_2, ..., s_N$ states, respectively (*e.g.*, if $N = 2$, $s_1 = 2$, and $s_2 = 3$, then x_1 is a binary random variable and x_2 is a ternary random variable). In general, how many independent parameters in terms of N and the s_i ($1 \leq i \leq N$) do you need to specify the joint probability $p(x_1, x_2, ..., x_N)$? Justify your answer. Why is this impractical, and how does it motivate the need for conditional independence?

Q2. (15 points) After painstakingly working out your decision tree, you decide, however, that the results are inconclusive and that you should consider other higher-stake factors, such as the "Hokie plague" (probably some kind of flu, but it remains a medical mystery) going around, as well as the very important assignment due the next day. Due to the high numbers of people carrying the plague, you know that **G**oing out right now may significantly increase your chance of contracting the **H**okie plague. The plague might make you **D**rowsy, although drowsiness could also be caused by a lack of **S**leep. You know that **I**nsomnia could either prevent you from getting good sleep, or make it hard for you to **P**ay attention. Being drowsy and not being able to pay attention could both result in you unable to complete your **A**ssignment.

1. (3 points) Using the bolded and underlined letters (as indicated above) to represent the random variables, construct the corresponding Bayesian network. Include your graph in this write-up. Write out the joint probability $p(G, H, D, S, I, P, A)$. Which nodes are in S's Markov blanket?

For the following questions, you are to determine whether these conditional independence statements are true or not, and what they intuitively mean accordingly.

2. (3 points) $S \perp\!\!\!\perp P \mid I$.
3. (3 points) $H \perp\!\!\!\perp I \mid A$.
4. (3 points) $G \perp\!\!\!\perp I \mid S$.
5. (3 points) $G \perp\!\!\!\perp I \mid S, A$.