Homework 3

Automated Learning and Data Analysis
Dr. Thomas Price

Spring 2021

Instructions

Due Date: April, 22 2021 at 11:45 PM
Total Points: 50 for CSC522; 45 for CSC422.
Submission checklist:

- Clearly list each team member's names and Unity IDs at the top of your submission.
- Your submission should be a single PDF file containing your answers. **Name your file**: G(homework group number)_HW(homework number), e.g. G1_HW3.
- If a question asks you to explain or justify your answer, **give a brief explanation** using your own ideas, not a reference to the textbook or an online source.
- Submit your PDF through Gradescope under the HW3 assignment (see instructions on Moodle). **Note**: Make sure to add you group members at the end of the upload process.
- In addition to your group submission, please also *individually* submit your Programming portion via our JupyterHub site *and* Moodle.

Problems

1. BN Inference (12 points) [**Chengyuan Liu**]. Compute the following probabilities according to the Bayesian net shown in Figure 1. **Note**: $P(A)$ means $P(A = true)$; $P(\sim A)$ means $P(A = false)$.

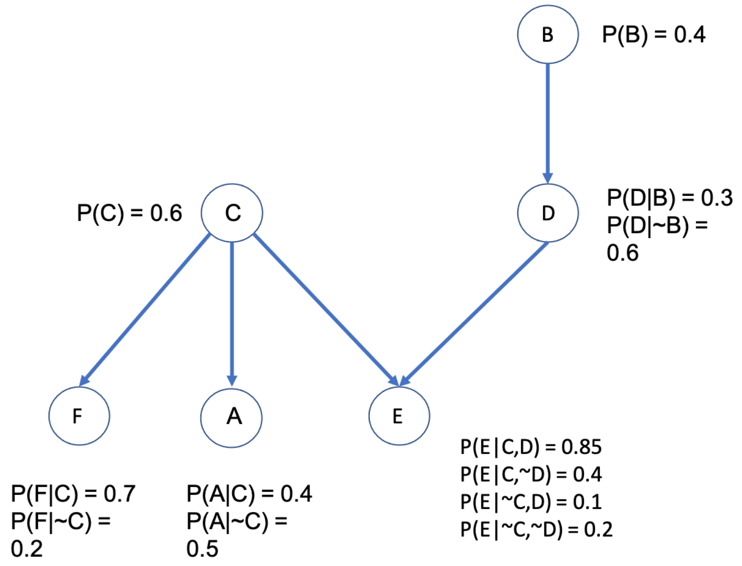

Figure 1: BN Inference

   (a) Compute $P(A)$. Show your work.
   (b) Compute $P(D|B, \sim A)$. Show your work.
   (c) Compute $P(A, B, \sim C, D, E, F)$. Show your work.
   (d) Are E and F conditionally independent given C? Justify your answer in 1 sentence.
   (e) Are A and B marginally independent? Justify your answer in 1 sentence.
   (f) Given evidence that $A = true$, $C = true$ $D = false$, and $F = true$, use the Bayes Net to predict whether E is more likely to be $true$ or $false$, or whether both are equally likely.

2. Linear Regression (18 points) [**Chengyuan Liu**].

   (a) Given the following four training data points of the form (x, y): $(4, 5)$, $(0, -2)$, $(1, -3)$, $(9, -4)$, estimate the parameters for linear regression of the form $y = w_1 x^{0.5} + w_0$.
   Note that we use the square root of x in the formula.
      i. Determine the values of w_1 and w_0 and show each step of your work.
      ii. Calculate the training RMSE for the fitted linear regression.

3. ANN + Backpropagation (20 pts (522 students) / 15 pts (422 students)) [**Chengyuan Liu**].

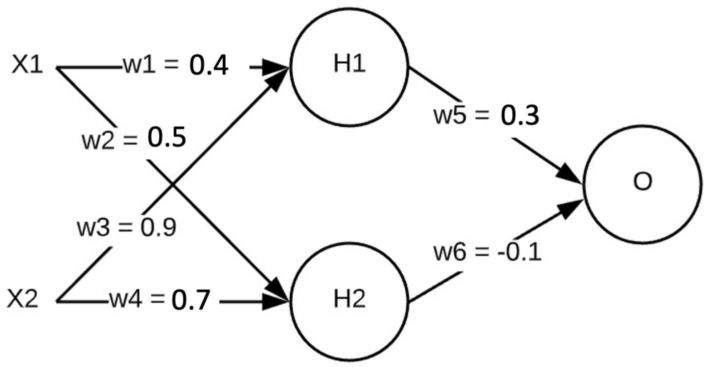

Figure 2: Neural Network Structure with initial weights

Table 1: Initial weights for given neural network in (a)

Weight	From	To	Initial Value
w1	X1	H1	0.4
w2	X1	H2	0.5
w3	X2	H1	0.9
w4	X2	H2	0.7
w5	H1	O	0.3
w6	H2	O	-0.1

You are given the above (Figure 2) neural network with continuous input attributes $X1$ and $X2$ and continuous output variable Y. For clarity, the relationship between weights and activations is also shown in Table 1. All three activations $H1$, $H2$ and O use the linear activation function $f(z) = Mz$, with constant $M = 1$. Initial weights are as given in Figure 2 and repeated in Table 1. There is **no bias** (w_0) added to any of the units. Answer the following:

(a) **Forward Pass**: If you are given one training data point: $X1_i = 1$, $X2_i = -1$, and $Y_i = 1$. Compute the activations of the neurons $H1$, $H2$ and O.

(b) **Backward Pass**: At the end of forward pass, using the current training instance i: $X1_i = 1$, $X2_i = -1$, and $Y_i = 1$, calculate the updated value of each of the following weights after one iteration of backpropagation:

- For CSC 522: $w1$, $w5$ and $w6$
- For CSC 422: $w5$ and $w6$ ($w1$ is optional extra credit)

Use 0.1 as your learning rate and MSE (mean squared error) as your cost function. Show your work on the following steps for each weight, w ($w1$, $w5$, $w6$):

i. Consider only the training instance i. Let a_N be the activation at neuron N, $X1_i$ be the value of the attribute $X1$ for instance i, and Y_i be the actual class of the instance i. Write equations to define the following:

A. The cost function C in terms of Y_i and a_O (Since we are considering a single instance, you do not have to sum over instances.)

B. The activation of the final layer a_O in terms of second layer weights w_5, w_6 and the activation of the first layer a_{H1} and a_{H2}

C. The activation of the node a_{H1} in terms of inputs $X1_i$, $X2_i$ and weights w_1 and w_3

ii. For layer-2 weights, calculate $\frac{\delta C}{\delta a_O}$ and $\frac{\delta a_O}{\delta w}$. Here C is the cost function, a_O is the activation at node O, and w is the weight.

iii. For layer-1 weights, calculate $\frac{\delta C}{\delta a_O}$, $\frac{\delta a_O}{\delta a_{H1}}$, and $\frac{\delta a_{H1}}{\delta w}$ (522/bonus only).

iv. Calculate $\frac{\delta C}{\delta w}$ using the above values.

v. Calculate the updated weight w' using the $\frac{\delta C}{\delta w}$ and the learning rate.