Homework 3

Automated Learning and Data Analysis
Dr. Thomas Price

Spring 2022

Instructions

Due Date: April, 4 2022 at 11:45 PM
Total Points: 50 for CSC 522; 45 for CSC 422
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.
- Submit the programming portion of the homework *individually* through JupyterHub.

1 BN Inference (12 points) [Chengyuan Liu]

Compute the following probabilities according to the Bayesian net shown in Figure 1 (under the Causal Markov Assumption). **Note**: $P(A)$ means $P(A = true)$; $P(\sim A)$ means $P(A = false)$.

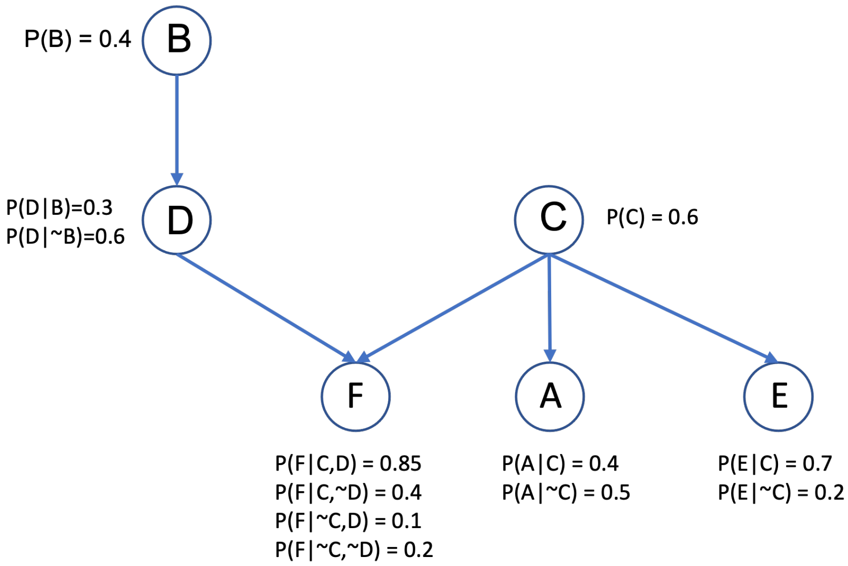

Figure 1: BN Inference

1a) (*2 points*) Compute $P(A)$. Show your work.

1b) (*2 points*) Compute $P(D|B, \sim A)$. Show your work.

1c) (*2 points*) Compute $P(A, B, \sim C, D, E, F)$. Show your work.

1d) (*2 points*) Under the Causal Markov Assumption, are E and F conditionally independent given C? Justify your answer in 1 sentence.

1e) (*2 points*) Under the Causal Markov Assumption, are A and B marginally independent? Justify your answer in 1 sentence.

1f) (*2 points*) Given evidence that $A = true$, $C = true$ $D = false$, and $E = true$, use the Bayes Net to predict whether F is more likely to be $true$ or $false$, or whether both are equally likely.

Solution:

1a) $P(A) = P(A|C)P(C) + P(A| \sim C)P(\sim C) = 0.4 * 0.6 + 0.5 * 0.4 = 0.44$

1b) $P(D|B, \sim A) = P(D|B) = 0.3$

1c) $P(A, B, \sim C, D, E, F) = P(A| \sim C)P(B)P(\sim C)P(D|B)P(E| \sim C)P(F| \sim C, D) = 0.5 * 0.4 * 0.4 * 0.3 * 0.2 * 0.1 = 0.00048$

1d) T. When C is given precisely, the occurrence of E and the occurrence of F are independent events. If C is unknown, E and F are not independent to each other.

1e) T. Knowledge of A's value doesn't affect your belief in the value of B and vice versa.

1f) $P(F|A, C, \sim D, E) = P(F|C, \sim D) = 0.4$
$P(\sim F|A, C, \sim D, E) = P(\sim F|C, \sim D) = 0.6$
Therefore, F is more likely to be false.

2 Linear Regression (18 points) [Benyamin Tabarsi]

2a) Given the following four training data points of the form (x, y): $(-2, 0)$, $(0, -1)$, $(2, -4)$, $(1, 2)$, estimate the parameters for linear regression of the form $y = w_1 x^2 + w_0$.
Note that we use the square of x in the formula.
Also report your answer to 2 decimal places (hundredths place).

i) (*14 points*) Determine the values of w_1 and w_0 and show each step of your work.

ii) (*4 points*) Calculate the training RMSE for the fitted linear regression. Show your work.

Solution:

- Solve the linear regression using least squares minimization:

$$\begin{aligned} SSE &= \sum_{i=1}^{4}(y_i - (w_1 x_i^2 + w_0))^2 \\ &= (-4w_1 - w_0)^2 + (-1 - w_0)^2 + (-4 - 4w_1 - w_0)^2 + (2 - w_1 - w_0)^2 \\ &= 4w_0^2 + 33w_1^2 + 18w_0 w_1 + 6w_0 + 28w_1 + 21 \end{aligned}$$

  Take derivatives of w_0 and w_1 and set them to 0:
  $\frac{\partial SSE}{\partial w_0} = 8w_0 + 18w_1 + 6 = 0$
  $\frac{\partial SSE}{\partial w_1} = 66w_1 + 18w_0 + 28 = 0$

  Solve the above equations and get the values of w_0 and w_1 as:
  $w_0 = 0.53$, $w_1 = -0.56$

- $y = [0, -1, -4, 2]$
  $predicated\_y = [-1.75, 0.53, -1.75, -0.04]$
  RMSE: 1.91

3 ANN + Backpropagation (20 points) [Jianxun Wang]

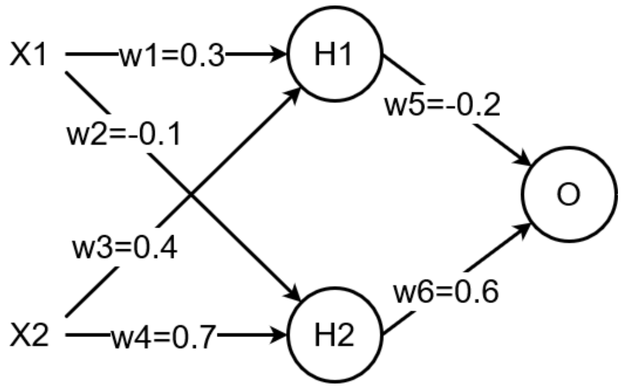

Figure 2: Neural Network Structure with initial weights

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

Weight	From	To	Initial Value
w1	X1	H1	0.3
w2	X1	H2	-0.1
w3	X2	H1	0.4
w4	X2	H2	0.7
w5	H1	O	-0.2
w6	H2	O	0.6

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. Calculation should be done with **4 decimal points**.

3a) (*3 points*) **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.

Solution:

Table 2: Output of each units for $X1 = 1$, $X2 = -1$, and $Y = 1$

Weight	Computed Value
a_{H1}	$= (0.3 - 0.4) = -0.1$
a_{H2}	$= (-0.1 - 0.7) = -0.8$
a_O	$= (-0.2 * -0.1 + 0.6 * -0.8) = -0.46$

3b) **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) (*3 points*) 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:

i) 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.)

ii) 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}

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

ii) (*2 points*) For layer-2 weights w, 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) (*4 points*) For layer-2 weights w, calculate $\frac{\delta C}{\delta w}$ of corresponding weights using the above values.

iv) (*3 points*) For $w1$, calculate $\frac{\delta C}{\delta a_O}$, $\frac{\delta a_O}{\delta a_{H1}}$, and $\frac{\delta a_{H1}}{\delta w1}$ **(522 only / 422 bonus)**.

v) (*2 points*) For $w1$, calculate $\frac{\delta C}{\delta w1}$ of corresponding weights using the above values **(522 only / 422 bonus)**.

vi) (*3 points*) For all weights, calculate the updated weight w' using the $\frac{\delta C}{\delta w}$ and the learning rate.

Solution:

$$C = (Y_i - a_O)^2$$

$$a_O = w5a_{H1} + w6a_{H2}$$

$$a_{H1} = w1X1 + w3X2$$

For the second layer weights,

$$\frac{\delta C}{\delta a_O} = -2(Y_i - a_O) = -2(1 - (-0.46)) = -2.92$$

$$\frac{\delta a_O}{\delta w5} = a_{H1} = -0.1$$

$$\frac{\delta a_O}{\delta w6} = a_{H2} = -0.8$$

Thus, the gradients for $w5$ and $w6$ are:

$$\frac{\delta C}{\delta w5} = \frac{\delta C}{\delta a_O}\frac{\delta a_O}{\delta w5} = -2.92 \times -0.1 = 0.292$$

$$\frac{\delta C}{\delta w6} = \frac{\delta C}{\delta a_O}\frac{\delta a_O}{\delta w6} = -2.92 \times -0.8 = 2.336$$

CSC 522 (extra for 422): For the second layer weight $w1$,

$$\frac{\delta a_O}{\delta a_{H1}} = w5 = -0.2$$

$$\frac{\delta a_{H1}}{\delta w1} = X1 = 1$$

Thus, the gradient for $w1$ is:

$$\frac{\delta C}{\delta w1} = \frac{\delta C}{\delta a_O}\frac{\delta a_O}{\delta a_{H1}}\frac{\delta a_{H1}}{\delta w1} = -2.92 \times -0.2 \times 1 = 0.584$$

Updating the weights:

$$w5' = w5 - \alpha\frac{\delta C}{\delta w5} = -0.2 - 0.1 \times 0.292 = -0.2292$$

$$w6' = w6 - \alpha\frac{\delta C}{\delta w6} = 0.6 - 0.1 \times 2.336 = 0.3664$$

(522 only/ 422 extra)

$$w1' = w1 - \alpha\frac{\delta C}{\delta w1} = 0.3 - 0.1 \times 0.584 = 0.2416$$