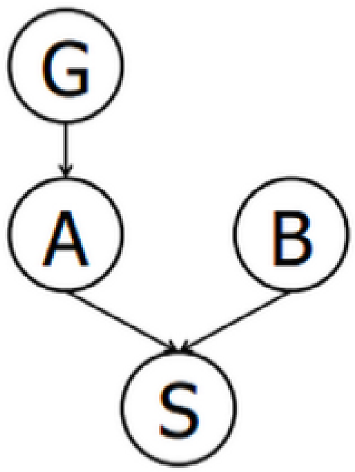
Q1 Disease Diagnosis
3 Points

Suppose that a patient can have a symptom (S) that can be caused by two different diseases (A and B). It is known that the variation of gene G plays a big role in the manifestation of disease A. The Bayes' Net and corresponding probability tables for this situation are shown below.

G	$P(G)$
g	0.10
$\neg g$	0.90

G	A	$P(A \mid G)$
g	a	1.00
g	$\neg a$	0.00
$\neg g$	a	0.10
$\neg g$	$\neg a$	0.90



B	$P(B)$
b	0.40
$\neg b$	0.60

A	B	S	$P(S \mid A, B)$
a	b	s	1.00
a	b	$\neg s$	0.00
a	$\neg b$	s	0.90
a	$\neg b$	$\neg s$	0.10
$\neg a$	b	s	0.80
$\neg a$	b	$\neg s$	0.20
$\neg a$	$\neg b$	s	0.10
$\neg a$	$\neg b$	$\neg s$	0.90

Round all answers to three decimal places.

Q1.1 P(g,a,b,s)
1 Point

Compute P (g, a, b, s), where g denotes that the variation of gene G is present, and a, b, s indicate that the diseases A and B and symptoms s are all present.

.04

Q1.2 P(g|a)
1 Point

What is the probability that a patient has the disease carrying gene variation G given that they have disease A?

.526

Q1.3 p(g|b)
1 Point

What is the probability that a patient has the disease carrying gene variation G given that they have disease B?

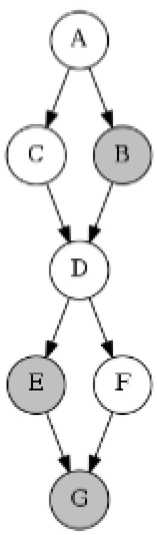
.1

Q2 Independence
4 Points

Answer the following questions about the Bayes nets shown below. For each, answer the yes/no question and give a brief explanation.

Q2.1 A⊥D|B,E,G
2 Points

In the Bayesian Network below, is it guaranteed that A is independent of D given E, B and G?



Yes

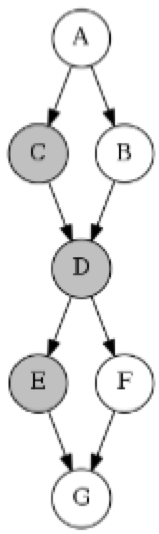
No

Why?

There doesn't seem to be a causal chain which guarantees that A is independent of D given E, B, and G. P(a,c,d,f,g) would cause it to not be independent. There also isnt anything which is blocking.

Q2.2 G⊥B|C,D,E
2 Points

In the Bayesian Network below, is it guaranteed that G is independent of B given C, D and E?



Yes

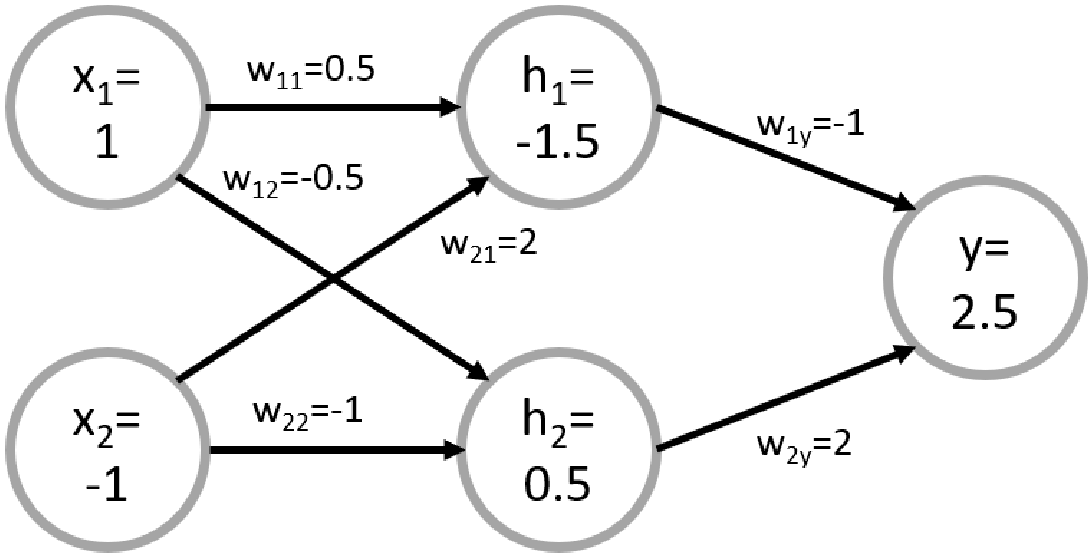
No

Why?

Given that c, d, and e are true there seems to be a guarantee that G is independent of B. We can see that there is no path that the net can take from B to G without first accessing D. this would cause independence. Also since D blocks B then G and B are independent.

Q3 Deep neural networks
3 Points

Consider the neural network shown below. There are no biases, and there is no non-linearity at the hidden layer.



Q3.1 Output
1 Point

There are two weights on the output layer: w_{1y} and w_{2y}. If the desired prediction is $y = 0$, which of these output weights would gradient descent try to INCREASE in this step?

w_{1y} only

w_{2y} only

Both

Neither

Q3.2 To h1
1 Point

In the same setup, which of the weights pointing to h_1 would gradient descent try to INCREASE?

w_{11} only

w_{21} only

Both

Neither

Q3.3 To h2
1 Point

In the same setup, which of the weights pointing to h_2 would gradient descent try to INCREASE?

w_{12} only

w_{22} only

Both

Neither

Q4 Deep network
5 Points

Consider the MNIST digit classification problem. Here, there are 784 input units (corresponding to pixels in a 28x28 grid). There are ten output units (corresponding to the digits 0 through 9).

Q4.1 Linear
1 Point

Suppose we trained a linear model WITH NO BIASES, that is "fully connected" - there is a weight connecting each input unit to each output unit. How many weights (=parameters) are there to estimate in this model? (Please write out the full number.)

7840

Q4.2 One hidden layer
1 Point

Suppose we introduced a single hidden layer with 100 hidden units. We fully connect the input to this hidden layer, and fully connect the hidden layer to the output layer. Again, no biases anywhere. How many weights (=parameters) are there to estimate in this model? (Please write out the full number.)

79400

Q4.3 Ten hidden layers

1 Point

Suppose instead of one hidden layer, we have ten hidden layers, each with 100 hidden units. Each hidden layer is fully connected to the one that comes before it and the one that comes after it. Again, no biases anywhere. How many weights (=parameters) are there to estimate in this model? (Please write out the full number.)

794000

Q4.4 Vanishing gradient
2 Points

What is the vanishing gradient problem, and why is it a problem?

the problem is as more layers using certain activation functions are added to neural networks, the gradients of the loss function approaches zero, making the network hard to train. This is a problem because certain activation functions, like the sigmoid function makes thresholding easier as it pushes large inputs into 1s and small inputs into 0s. this will cause a large change in the inputs to result in a large change in the outputs.

HW5

● Ungraded

Student
Alessandro Gagarin

Total Points
- / 15 pts

Question 1
Disease Diagnosis 3 pts

1.1	P(g,a,b,s)	1 pt
1.2	P(g|a)	1 pt
1.3	p(g|b)	1 pt

Question 2
Independence 4 pts

2.1 $A \perp D | B, E, G$ 2 pts

2.2 $G \perp B | C, D, E$ 2 pts

Question 3
Deep neural networks 3 pts

3.1 Output 1 pt

3.2 To h1 1 pt

3.3 To h2 1 pt

Question 4
Deep network 5 pts

4.1 Linear 1 pt

4.2 One hidden layer 1 pt

4.3 Ten hidden layers 1 pt

4.4 Vanishing gradient 2 pts