

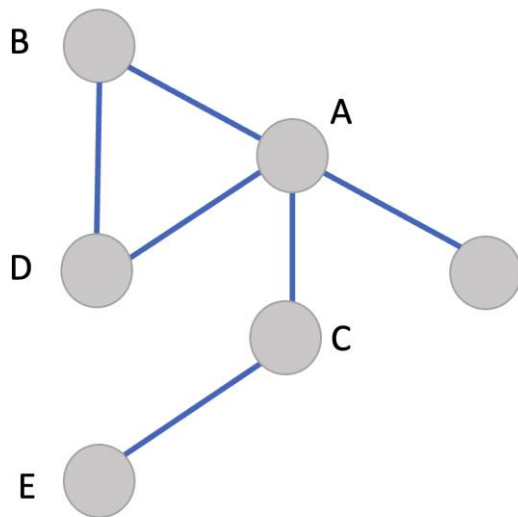
Name:

Netid:

Course section:

Discussion section:

1. Given the following network, answer the questions: (10 pt in total, 2pt each)



(a) The degree of node A?

4

(b) The clustering coefficient of node A?

1/6

(c) The clustering coefficient of node B?

1

(d) The clustering coefficient of node C?

0

(e) The shortest path length from B to E?

3

2. What is a scale-free network? Explain what preferential attachment is when you generate a scale-free network. (5 pt)

(2 pt) The distribution of node linkages follows a power law. there are a few highly connected nodes, called hubs, and many nodes with only a few connections.

-1 pt if not mention power law.

(3 pt) Preferential attachment is a mechanism by which nodes in a scale-free network tend to connect to other nodes that are already highly connected.

OR: "The rich get richer"

Quiz 2 CBB 752

3. For each of the following, choose whether they are examples of data privacy or data security: (8 pt in total, 2 pt each)

(a) An individual has the right to access their personal data gathered from lab tests, as well as the right to request that the data be deleted or corrected.

privacy

(b) A hospital uses firewalls to protect its servers from cyber attacks and unauthorized access to sensitive data.

security

(c) A blockchain ledger is used to protect the integrity of medical records, protecting it from either intentional or accidental corruption.

security

(d) A person may choose to share their genome with a company for a certain purpose, such as finding relatives, but they may not want that information to be used for other purposes, such as targeted advertising

privacy

4. The following equation describes a perceptron. x is the input, and z is the linear transformation of x . y is the output of the activation. Derive the gradient of y with respect to the weight w . (10 pt)

$$z=wx+b$$

$$y = \frac{1}{1+e^{-z}}$$

The image shows a handwritten derivation of the gradient of y with respect to w . The steps are as follows:

$$\begin{aligned} \frac{dy}{dw} &= \frac{dy}{dz} \cdot \frac{dz}{dw} \\ &= \frac{d}{dz} (1+e^{-z})^{-1} \cdot x \\ &= \sigma(z)(1-\sigma(z)) \cdot x \quad \text{OR} \quad y(1-y)x \quad \text{OR} \quad \frac{e^{-z} \cdot x}{(1+e^{-z})^2} \end{aligned}$$

- + 2 pt for correctly applying chain rule
- + 6 pt for partially correct answer
- + 10 pt for fully correct

5. Describe the difference between an Autoencoder and a Variational Autoencoder. (5 pt)

Autoencoders aim to minimize the reconstruction error between the original data and its compressed representation, while VAEs optimize the ELBO by maximizing the likelihood of the observed data given the latent variables and minimizing the KL divergence between the prior and posterior distributions of the latent variables. Autoencoders are deterministic, VAE are stochastic/probabilistic. VAEs learn a probability distribution over latent space.

-1 pt for minor mistake
-4 pt for incorrect answer

6. Given the input and kernel, set stride = 1, padding = 0 for convolution, and max pooling with 2x2 filters and stride = 1, then manually complete the convolution step and the max pooling step of CNN. (10 pt total, 5 pt

1	1	1	0	0
0	1	1	1	0
0	0	1	1	1
0	0	1	1	0
0	1	1	1	1

Input

1	1	1
1	1	0
1	0	0

Kernel

each)

- (a) Filling the matrix for the output of the convolution step.

-1 pt for minor mistake
-4 pt for major mistakes

4	4	4
2	4	5
1	4	6

(b) Filling the matrix for the output of the max pooling step.

-1 pt for minor mistake

-4 pt for major mistakes

4	5
4	6

7. Explain the gradient exploding/vanishing problem and name one way to overcome this issue. (5 pt)

(+ 3 pt) The gradient exploding/vanishing problem occurs when gradients become either too large or too small during backpropagation in deep neural networks.

(+2 pt) Methods to address this issue include using alternative activation functions that have better gradient behavior, such as the rectified linear unit (ReLU), or using normalization techniques such as batch normalization, which can help stabilize the gradients during training.

-1 pt for minor mistakes

8. List 1 challenge in biosensor data processing, and 1 solution for the challenge. (5 pt)

presence of noise/ interference/ signal variability/non-specific binding

sensitivity

specificity

machine learning techniques

1 challenge +2pt

1 solution +3pt

9. Name two properties of intrinsically disordered proteins (IDPs) that distinguish them from folded proteins. (5 pt)

lack of stable tertiary structure/conformational flexibility

more charged

less hydrophobic

more likely to aggregate/phase separate

one properties +2pt

two properties +3pt

10. Name a property of the protein core (rather than surface residues) that is important to study and use for understanding protein structure and function. (5 pt)

packing density +5pt

hydrophobicity

11. Suppose a linear polymer is made up of 10 identical spherical atoms in three dimensions. How many bond angles and dihedral angles does this polymer possess? (5 pt)

8 bond angle +2pt

7 dihedral angle +3pt

12. What is the packing fraction of an all-atom hard-sphere (i.e. explicit hydrogen) representations of residues in protein cores? Name one reason for why residues in protein cores have a smaller packing fraction than the densest packing fraction for identical hard spheres. (5 pt)

0.74 +2pt

flexibility of protein backbone and sidechain/ have voids or channels/ not identical hard spheres +3pt

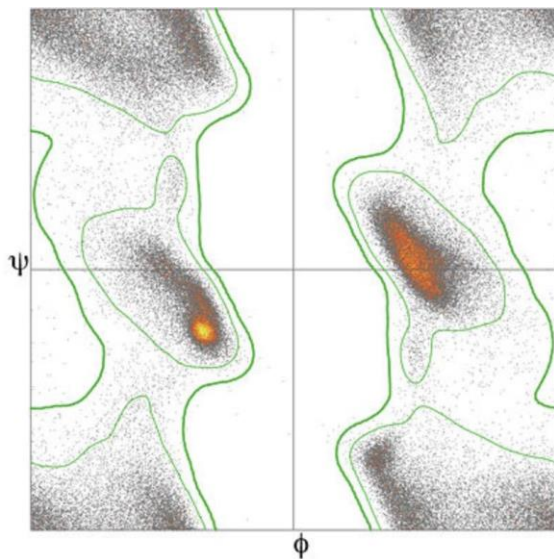
*****but also, can be 0.55-0.56 in reality (if only gave this number i took off -1pt) ideally they have both numbers and explain why
NOT on a lattice

13. What does it mean if a residue falls outside the allowed regions in a Ramachandran plot? (5 pt)

if wrong -5pt

- a) There may be a problem with the protein structure, such as a misaligned backbone or incorrect torsion angles
- b) The residue is likely to be involved in protein-protein interactions
- c) The residue is likely to be buried in the interior of the protein

14. **BONUS:** Below is a Ramachandran plot of one amino acid. Name which amino acid it is and why do you think so? (5 pt)



Glycine (no side chain, so more areas of accessible phi and psi)

amino acid correct +3pt

reason correct +2pt

15. **BONUS:** Describe what Alpha Fold does and why it improves over other protein folding approaches. Why type of models/architectures does Alpha Fold use compared to other models? (5 pt)

The architecture based on a combination of CNN and RNN (or transformers, AF2) +2pt

Accuracy or speed over protein folding approaches +1pt

Predict protein structure from amino acid sequence, also mention multiple sequence alignment +2pt