

1. Gradient

Derive the gradient of the cross-entropy loss function with respect to the input x

$$L = -t \log y - (1-t) \log(1-y)$$

Where t is the ground truth label, $y = Wx+b$

Notice: for the simplicity of calculation, there is no non-linear activation (8 pt)

$$\frac{\partial L}{\partial x} = \frac{\partial L}{\partial y} \frac{\partial y}{\partial x} = -\left(\frac{t}{y} - \frac{1-t}{1-y}\right)W$$

Each term: 3pt x2

Final answer: 2pt

Can accept if log is treated as log10

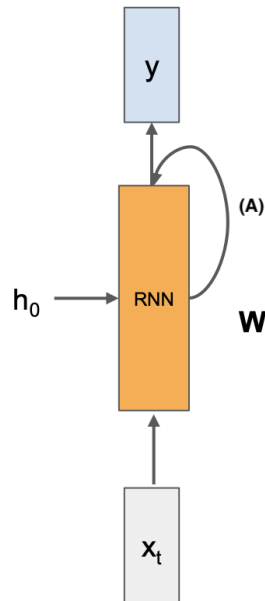
2. RNN

(1) Below is a recurrent neural network diagram. What does the arrow "A" indicate? (4 pt)

Hint: it relates to how the recurrent neural network is "recurrent"

The **value/output** of the hidden layer for one position is fed as the input to the next position

Just saying "data" is not sufficient. -2pt



(2) Explain the exploding gradient problem. Why is it a major problem in vanilla recurrent neural networks? (4pt x2)

Exploding gradient happens when many terms with values greater than 1 are multiplied.

The "term" here should be a **gradient or derivative**. If the answer specifies "weight" "loss" etc. -2pt

In RNN, if the sequence is long, (1) the network becomes very deep (2) the same term is repeated a lot of times (3) unbounded activation function may lead to exploding gradient (this is not exclusive to RNN but also acceptable) (either point is sufficient)

This answer is also accepted because the question is a little ambiguous:

It makes convergence difficult/training unstable, etc.

3. You're given a set of biomedical images to classify. Describe two ways you can augment the data. (5 pt x2)

Any kind of rotation/cropping/random masking/random noise, etc.

4. CNN

(1) Calculate the convolution result of the following matrix with the given kernel, padding = 1, stride = 2 (6 pt)

The matrix:

1	3
-1	2

The kernel:

1	-1
1	0

0	3
1	2

Each cell 1pt

Final answer 2pt

(2) Calculate the max pooling result of the given matrix with a 2x2 filter and stride=1: (6 pt)

1	0	4
5	2	7
3	5	1

5	7
5	7

Each cell 1pt

Final answer 2pt

5. True or false: (2pt x4)

- (1) Latent Dirichlet allocation (for topic modeling) is a supervised algorithm **False**
- (2) A typical GAN consists of a generator network and a discriminator network **True**
- (3) Policy gradient is a reinforcement learning method **True**
- (4) Variational autoencoder is a discriminative model **False**

6. Figure 1 below shows a polypeptide chain with four amino acids. The amino acids are labeled as follows: $i-1$, i , $i+1$, and $i+2$. The R-groups indicate the side chains that extend from the α atoms. Provide the sequences of atoms that define the Φ and Ψ backbone dihedral angles for amino acid i . Use the notation N^i , C_α^i and C^i for the nitrogen, α -Carbon, and carboxyl carbon on the i th amino acid. (5pt x2)

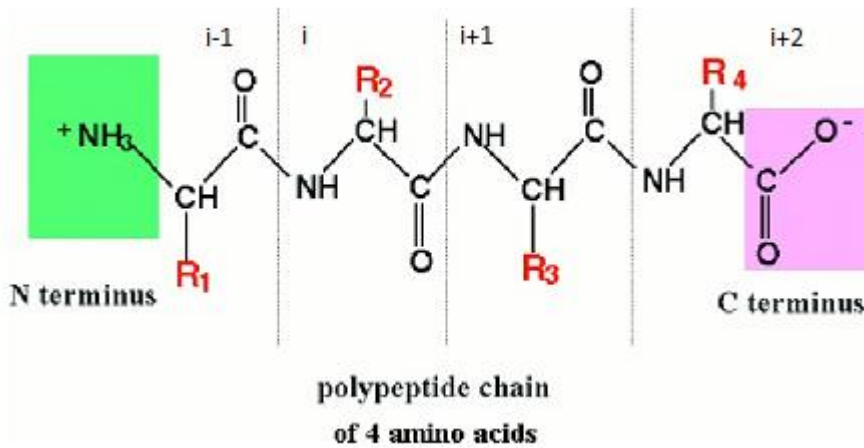


Figure 1: Schematic of a polypeptide chain with 4 amino acids.

Φ (Φ) $C^{i-1} N^i C_\alpha^i C^i$

Ψ (Ψ) $N^i C_\alpha^i C^i N^{i+1}$

7. Explain how the Voronoi volume of a residue for a protein is defined. (10 pt)

It is defined by the region of space that is closer to that residue than any other residue.

8. The densest packing fraction for identical hard spheres is approximately 0.74. Name one reason for why residues in protein cores have a smaller packing fraction than this. (10 pt)

Core residues are not ordered, do not exist in a lattice

Residues have rough surfaces

(only saying higher aspect ratio not correct)

9. Name two properties of intrinsically disordered proteins (IDPs) that distinguishes them from folded proteins (5 pt x2)

Intrinsically disordered proteins tend to be more electrically charged, and tend to be less hydrophobic than folded proteins. No well-defined tertiary structure/dynamic conformational changes, can liquid-liquid phase separate/aggregate

10. Suppose a linear polymer is made up of 5 identical spherical atoms in three dimensions as shown in Figure 2. How many bond angles and dihedral angles does this polymer possess? (5 pt x2)



Figure 2: Schematic of polymer with 5 monomers.

3 bond angles

2 dihedral angles

(Extra Credit) Figure 3 below shows a picture of a valine amino acid. Circle the four atoms that define the χ_1 side chain dihedral angle for valine. (10 pt)

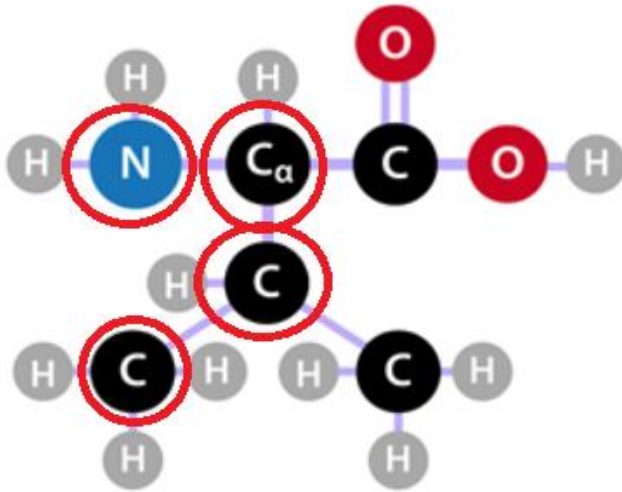


Figure 3: Schematic of valine residue.

(either methyl C is fine)

-5 points if carboxyl C or any H included

no points if the circled 4 atoms aren't in a sequence (not a dihedral angle)