

Name: \_\_\_\_\_

1. Given the following confusion matrix, select all the statements that accurately define sensitivity and specificity using TP, TN, FP, and FN. [5 points]

	Predicted Positive	Predicted Negative
True	TP	TN
False	FP	FN

- A. Sensitivity =  $TP / (TP + FN)$
  - B. Specificity =  $TN / (TN + FP)$
  - C. Sensitivity =  $TP / (TP + FP)$
  - D. Specificity =  $TN / (TN + FN)$
  - E. None of the above
2. Decision Tree: [5 points]  
 Imagine that you are a computer. Construct an optimal decision tree (with depth of 2) based on the following input data to predict patient survival. (Partial credit for any reasonable decision tree):

Tumor Degree	Tumor Size	Age	Patient Survival
I	10	40	Yes
II	10	60	Yes
I	30	50	No
I	13	40	No
II	10	50	No

Name: \_\_\_\_\_

3. Running the following SQL statement on the left generates the DB table on the right. Construct an SQL query that lists all patients whose age is 40 or above. [Optional for MCDB/MBB, 5 points]

```
CREATE DATABASE Patient_DB;  
  
CREATE TABLE Patient_DB.Patient  
(  
  ID int,  
  Name varchar (50),  
  Address varchar (250),  
  Age smallint  
  Sex varchar (2)  
);
```

ID	Name	Address	Age	Sex
1	John Doe	XYZ	40	M
2	Jane Smith	ABC	34	F
3	Mary Queen	PQSRT	46	F
4	Mike Lee	DWQER	60	M

4. Describe at least one of the differences between X-ray crystallography and NMR. [Optional for CBB/CPSC, 5 points]

5. Place the following NGS steps in the correct order. [5 points]

- A. Sequencing
- B. Library preparation
- C. Computational analysis
- D. Isolation of sample

Name: \_\_\_\_\_

6. The following steps describe an approach for discovering motifs using position weight matrix (PWM).

Step 1. Guess an initial weight matrix

Step 2. Use weight matrix to predict instances in the input sequences

Step 3. Use instances to predict a weight matrix

Step 4. Repeat 2 & 3 until satisfied

What is the name of algorithm used in this example? [5 points]

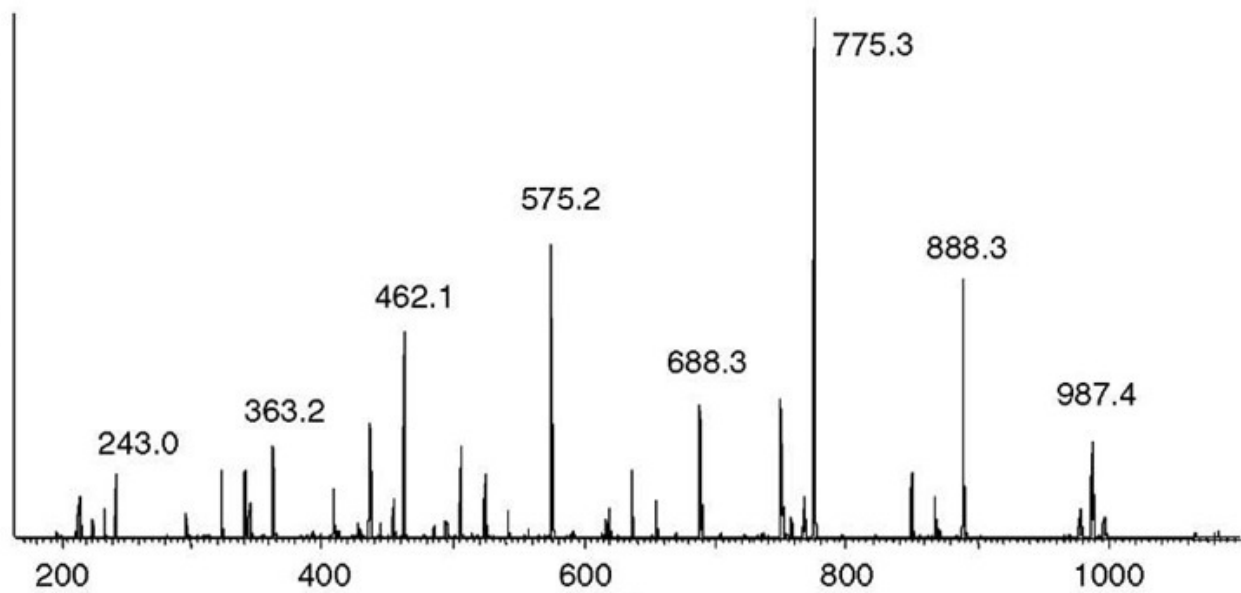
7. The first principal component produced by Principal Components Analysis (PCA) maximizes \_\_\_\_\_. [5 points]

8. What is the key difference between a PAM-50 and a PAM-500 substitution matrix? [5 points]

9. SILAC uses \_\_\_\_\_ to label cells grown in different conditions. (Hint: what does SILAC stand for?) [5 points]

Name: \_\_\_\_\_

10.



Mass spectrometry is a powerful technique to differentiate distinct proteins. A sample mass spectrum is shown above. What does X-axis and Y-axis of mass spectrum denote? [5 point]

X-axis:

Y-axis:

11. Illumina reads are generally 50nt - 250nt in length. What limits from producing longer read? [5 points]

12. List at least two things that can be studied with RNA-seq. [5 points]

Name: \_\_\_\_\_

13. List at least two limitations of ChIP-seq? [5 points]

14. Position Weight Matrix: [10 points]

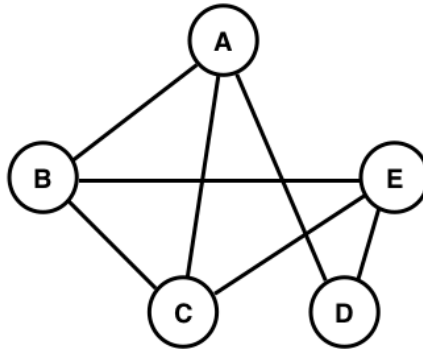
Given the following DNA sequences, complete the corresponding position probability matrix profile. Using the profile, calculate the probability of the sequence  $S = \text{GAGGT}$  being observed.

**DNA 1 : GAGGT**  
**DNA 2 : TCCGT**  
**DNA 3 : CAGGT**  
**DNA 4 : ACAGT**  
**DNA 5 : TAGGT**  
**DNA 6 : TAGGT**  
**DNA 7 : ATGGT**  
**DNA 8 : CAGGT**

	Pos 1	Pos 2	Pos 3	Pos 4	Pos 5
A	0.25				
C	0.25				
G	0.125				
T	0.375				

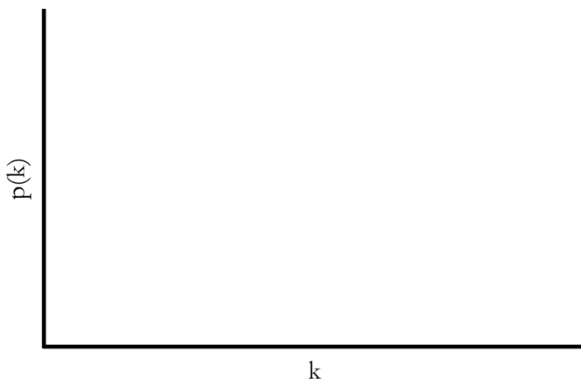
Name: \_\_\_\_\_

15. Calculate the average clustering coefficient of the following network. [5 points]

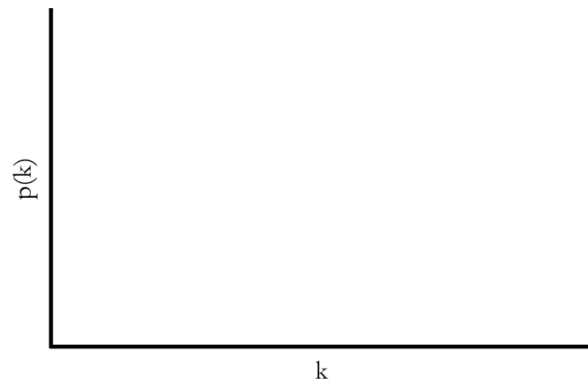


16. Given  $k$  is degree and  $p(k)$  is frequency, complete the degree distribution curves for random network and scale-free network. [5 points]

Random network



Scale-free network



Name: \_\_\_\_\_

17. Use the appropriate dynamic programming algorithm to LOCALLY align the sequences below. Use +2 for a match, -2 for a mismatch, and -1 for a gap. Complete the matrix shown below and show ALL optimal alignments. [20 points]

		<b>C</b>	<b>T</b>	<b>G</b>	<b>T</b>	<b>T</b>
<b>G</b>						
<b>C</b>						
<b>T</b>						
<b>G</b>						
<b>C</b>						