

Instructions: You are allowed to discuss but the final answer should be your own. Any instance of cheating will be considered as academic dishonesty and penalty will be applied.

1. **[6 Marks]** Compute the rank and the inverse of the following matrices, if it exists (; denotes the values in next row).
 (a) $\begin{bmatrix} 1 & 2 & 1 \\ 1 & 3 & 4 \\ 2 & 3 & -1 \end{bmatrix}$ (b) $\begin{bmatrix} 0 & -2 & 4 \\ 1 & 1 & -1 \\ 2 & 4 & -5 \end{bmatrix}$ and (c) $\begin{bmatrix} 1 & 2 & 1 & 0 \\ 2 & 5 & 5 & 1 \\ -2 & -3 & 0 & 3 \\ 3 & 4 & -2 & -3 \end{bmatrix}$
2. **[4 Marks]** There are five observations. You are given that their mean is 3. Sum of squares of the first 3 samples is 14. Sum of squares of the last 3 samples is 50. Sum of squares of the first two samples is 5. What will be the sample standard deviation?
3. **[10 Marks]** Short questions:
 - (a) When a decision tree is grown to full depth, it is more likely to fit the noise in the data **True/False**
 - (b) When the feature space is larger, over fitting is more likely. **True/False**
 - (c) Can you represent the following Boolean function with a single logistic threshold unit (i.e., a single unit from a neural network)? If yes, show the weights. If not, explain why not in 1-2 sentences.

A	B	F(A,B)
1	1	0
0	0	0
1	0	1
0	1	0

(d) Boolean random variables A and B have the joint distribution specified in the table below

A	B	F(A,B)
1	1	0.12
0	0	0.32
1	0	0.08
0	1	0.48

- (i) calculate: $P(A=0)$, $P(A=1)$, $P(B=0)$, and $P(B=1)$
 - (ii) Are A and B independent? Justify your answer.
4. **[5 Marks]** Given 6 data points in 5-d space, $(1, 1, 1, 0, 0)$, $(-3, -3, -3, 0, 0)$, $(2, 2, 2, 0, 0)$, $(0, 0, 0, -1, -1)$, $(0, 0, 0, 2, 2)$, $(0, 0, 0, -1, -1)$. We can represent these data points by a 6×5 matrix X, where each row corresponds to a data point.
 - (a) What is SVD of the data matrix X you choose?
 - (b) What is first principle component for the original data points?
5. **[15 Marks]** For each of the supervised learning methods that we have studied, indicate how the method could overfit the training data (consider both your design choices as well as the training) and what you can do to minimize this possibility. There may be more than one mechanism for overfitting, make sure that you identify them all.
 - A. Nearest Neighbors
 - i. How does it overfit?
 - ii. How can you reduce overfitting?
 - B. Decision Trees
 - i. How does it overfit?

ii. How can you reduce overfitting?

6. [10 Marks] For the following network find the best weight and bias parameters using the training samples and weights provided below:

Id	p1	p2	p3	p4	Class
1	252	4	155	175	1
2	175	10	186	200	1
3	82	131	230	100	0
4	115	138	80	88	0

$$W^1 = [w_{11} \ w_{12}; w_{21} \ w_{22}; w_{31} \ w_{32}; w_{41} \ w_{42}; w_{51} \ w_{52}]$$

$$= [-0.00469 \ 0.00797; -0.00256 \ 0.00889; 0.00146 \ 0.00322; 0.00816 \ 0.00258; -0.00597 \ -0.00876]$$

$$W^2 = [w_{11} \ w_{12}; w_{21} \ w_{22}; w_{31} \ w_{32}]$$

$$= [-0.00588 \ -0.00232; -0.00647 \ 0.00540; 0.00374 \ -0.00005]$$

