

## Homework-4

### Question 1

The following table consists of training data from an employee database.

department	status	age	salary
sales	senior	31-40	Medium
sales	junior	21-30	Low
sales	junior	31-40	Low
systems	junior	21-30	Medium
systems	senior	31-40	High
systems	junior	21-30	Medium
systems	senior	41-50	High
marketing	senior	31-40	Medium
marketing	junior	31-40	Medium
secretary	senior	41-50	Medium
secretary	junior	21-30	Low

Given an instance with the values: systems, senior, and 21-30 for the attributes department, status, and age, respectively, what would be a naive bayesian classification for the salary of the sample?

### Question 2

You are given the following training data.

$x$	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
label	A	A	A	A	B	A	A	A	A	B	B	B	B	A	B	B	B	B

1. What would be the classification of a test sample with  $x = 4.2$  according to 1-NN ?

**Answer:**  $A / B$

2. What would be the classification of a test sample with  $x = 4.2$  according to 2-NN ?

**Answer:**  $A / B$

3. What would be the classification of a test sample with  $x = 4.2$  according to 3-NN ?

**Answer:**  $A / B$

4 Use “leave-one-out” cross validation to estimate the error of 1-NN. If you need to choose between two or more examples of identical distance, make your choice so that the number of errors is maximized.

**Answer:**  $\frac{?}{18}$

5 Use “leave-one-out” cross validation to estimate the error of 2-NN. Whenever you need to make a choice between equal distance data or determining a majority, make your choice so that the number of errors is maximized.

**Answer:**  $\frac{?}{18}$

6 Use “leave-one-out” cross validation to estimate the error of 3-NN. Whenever you need to make a choice between equal distance data or determining a majority, make your choice so that the number of errors is maximized.

**Answer:**  $\frac{?}{18}$

- 7 Use “leave-one-out” cross validation to estimate the error of 4-NN. Whenever you need to make a choice between equal distance data or determining a majority, make your choice so that the number of errors is maximized.

**Answer:**  $\frac{?}{18}$

- 8 Use “leave-one-out” cross validation to estimate the error of 17-NN. Whenever you need to make a choice between equal distance data or determining a majority, make your choice so that the number of errors is maximized.

**Answer:**  $\frac{?}{18}$

### Question 3

Consider the following training data:

$x_1$	$x_2$	$y$
1	1	+
2	1	+
1	2	+
0	0	−
1	0	−
2	0	−
3	0	−
0	3	−
3	3	−

Here is an illustration of the data as 2D points:

−	−	−	−
·	+	+	·
·	+	·	·
−	·	·	−

1. Assume Gaussian distribution where both covariance matrices are a multiple of the identity matrix (Case 1.). What is the discriminat function?
2. Assume Gaussian distribution where the covariance matrix is the same for both classes (Case 2.). What is the discriminat function?
3. Assume equal priors and the most general Gaussian distribution (Case 3). What is the discriminat function?