1

Marks per question : 10 Marks Scored : 8.0

Q No.	Q. Type	Status	Marks	Comment	
1	File Upload	✔	8.0	-	Hide Answer

Using the candidate elimination algorithm on the given dataset below, find the learned version space.

Instances X:

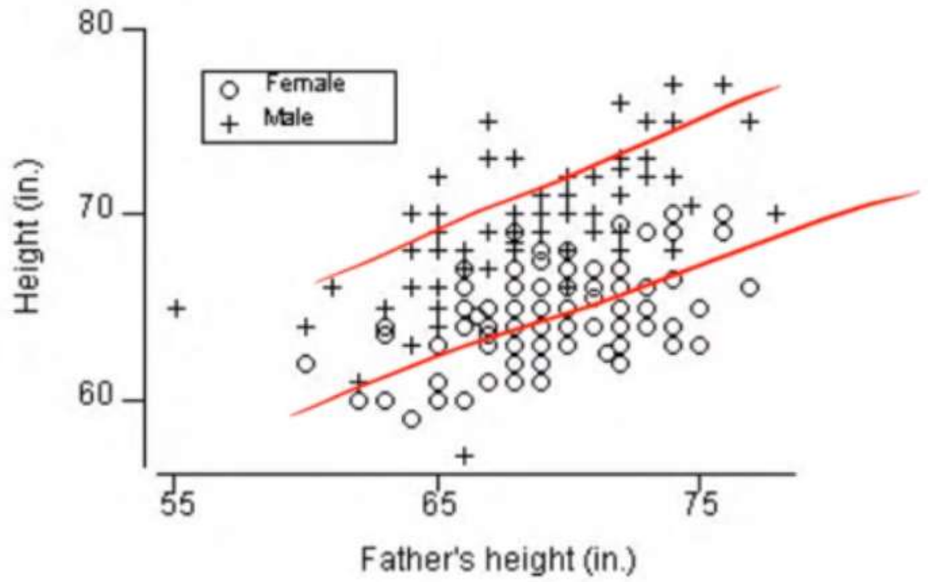
- *Domain*: edu, com, org
- *Platform*: Mac, PC, Unix
- *Browser*: Netscape2, Netscape 3, Netscape Communicator, Microsoft IE.
- *Day:* Monday - Sunday.
- *Screen:* VGA or XVGA.
- *Continent:* America, Europe, Africa, Asia, Australia.

Dom.	Plat.	Browser	Day	Screen	Cont.	Click?
edu	Mac	Net3	Mon.	XVGA	America	Yes
com	Mac	NetCom	Tue.	XVGA	America	Yes
com	PC	IE	Sat.	VGA	Eur.	No
org	Unix	Net2	Wed.	XVGA	America	Yes

The following table contains dataset consists of factors that affecting sunburn Use ID3 algorithm to construct a minimal decision tree that predicts whether a person is affected by sunburn or not. **SHOW EACH STEP OF THE COMPUTATION**

Name	*Hair*	*Height*	*Weight*	*Lotion*	*Result*
Sarah	blonde	average	light	no	sunburned (positive)
Dana	blonde	tall	average	yes	none (negative)
Alex	brown	short	average	yes	none(negative)
Annie	blonde	short	average	no	sunburned(positive)
Emily	red	average	heavy	no	sunburned(positive)
Pete	brown	tall	heavy	no	none(negative)
John	brown	average	heavy	no	none(negative)
Katie	blonde	short	light	yes	none(negative)

The scatterplot below shows student heights (y axis) versus father's heights (x axis) for a sample of 173 college students. The symbol "+" represents a male student and the symbol "o" is represents a female student.



1. Based on the scatterplot, what is the problem with using a regression equation for all 173 students?
2. The regression equation between father's heights and student's heights for the 79 male students is $\hat{y} = 30 + .58\,x$. One student was 72 inches tall; his father's height was 65 inches. What is the estimated height for this student,
3. What is the residual(Error) in problem 2 ?

Explain Logistic regression with multi class classification. How its cost function is different from binary classification.

Consider the following training data with features X1, X2 and Y is the outcome.

Data

X1	X2	Y
3	4	+
5	2	+
6	4	+
4	3	+
2	5	+
5	4	+
6	6	+
6	7	+
2	2	+
1	5	+
4	9	-
4	7	-
6	10	-
4	8	-
8	4	-
7	6	-
6	10	-
8	7	-
7	5	-
4	6	-
8	6	-
5	8	-
6	6	-
9	5	-
4	6	?

Use K-NN algorithm and find the class of (x1 = 4, x2=6) with K = 3,5,7 and 9 (Use Euclidian distance)

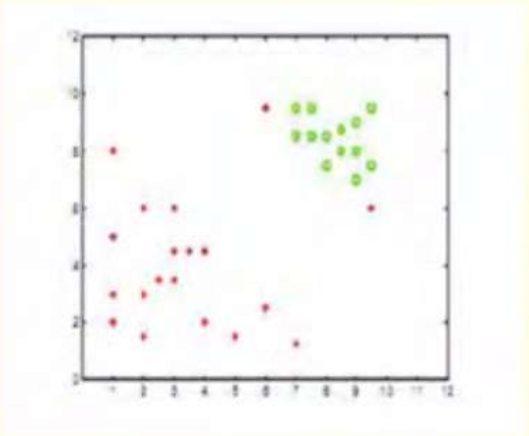
The table below lists a dataset that was used to create a nearest neighbour model that predicts whether it will be a good day to go surfing.

ID	Wave Size (ft)	Wave Period (secs)	Wind Speed (mph)	Good Surf
1	6	15	5	yes
2	1	6	9	no
3	7	10	4	yes
4	7	12	3	yes
5	2	2	10	no
6	10	2	20	no

Assuming that the model uses Euclidean distance to find the nearest neighbour, what prediction will the model return for each of the following query instances.

ID	Wave Size (ft)	Wave Period (secs)	Wind Speed (mph)	Good Surf
Q1	8	15	2	?
Q2	8	2	18	?
Q3	6	11	4	?

The original SVM proposed was a linear classier. In order to make SVM non-linear we map the training data on to a higher dimensional feature space and then use a linear classier in the that space. This mapping can be done with the help of kernel functions.
For this question assume that we are training an SVM with a quadratic kernel - i.e. our kernel function is a polynomial kernel of degree 2. This means the resulting decision boundary in the original feature space may be parabolic in nature. The dataset on which we are training is given below:



The slack penalty C will determine the location of the separating parabola. Please answer the following questions qualitatively.

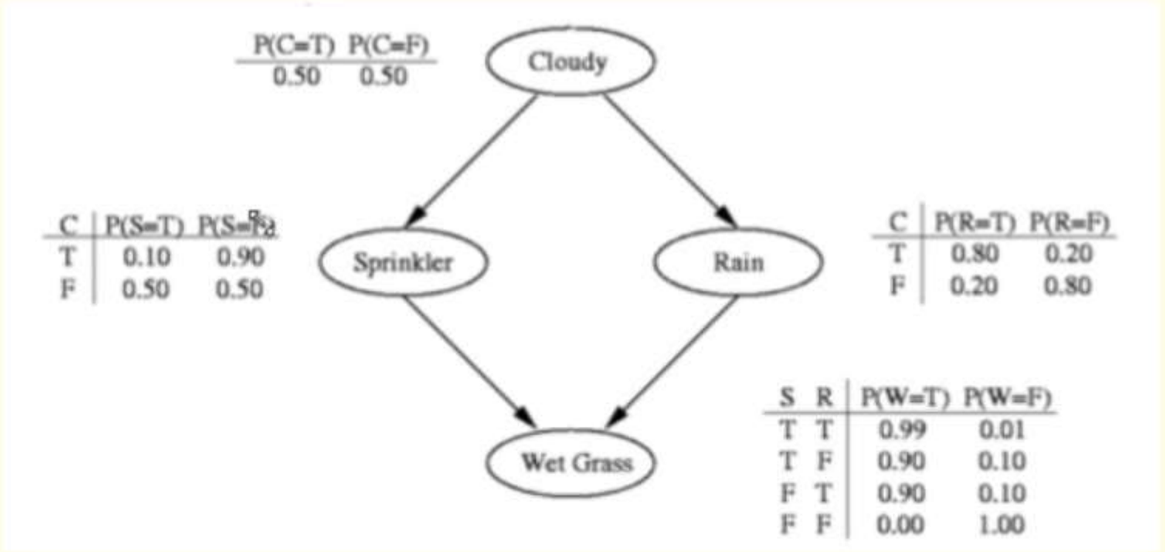
1. Where would the decision boundary be for very large values of C? (Remember that we are using a quadratic kernel). Justify your answer in one sentence and then draw the decision boundary
2. Where would the decision boundary be for C nearly equal to 0? Justify your answer
in one sentence and then draw the decision boundary
3. Now suppose we add three more data points as shown in figure below. Now the
data are not quadratically separable, therefore we decide to use a degree-5 kernel and find the following decision boundary. Most probably, our SVM suffers from a phenomenon which will cause wrong classification of new data points. Name that phenomenon, and in one sentence, explain what it is.

The table below gives details of symptoms that patients presented and whether they were suffering from meningitis.

ID	HEADACHE	FEVER	VOMITING	MENINGITIS
1	true	true	false	false
2	false	true	false	false
3	true	false	true	false
4	true	false	true	false
5	false	true	false	true
6	true	false	true	false
7	true	false	true	false
8	true	false	true	true
9	false	true	false	false
10	true	false	true	true

Using this dataset calculate the following probabilities:

a. $P(\text{VOMITING} = \textit{true})$
b. $P(\text{HEADACHE} = \textit{false})$
c. $P(\text{HEADACHE} = \textit{true}, \text{VOMITING} = \textit{false})$
d. $P(\text{VOMITING} = \textit{false} \mid \text{HEADACHE} = \textit{true})$
e. $P(\text{MENINGITIS} \mid \text{FEVER} = \textit{true}, \text{VOMITING} = \textit{false})$

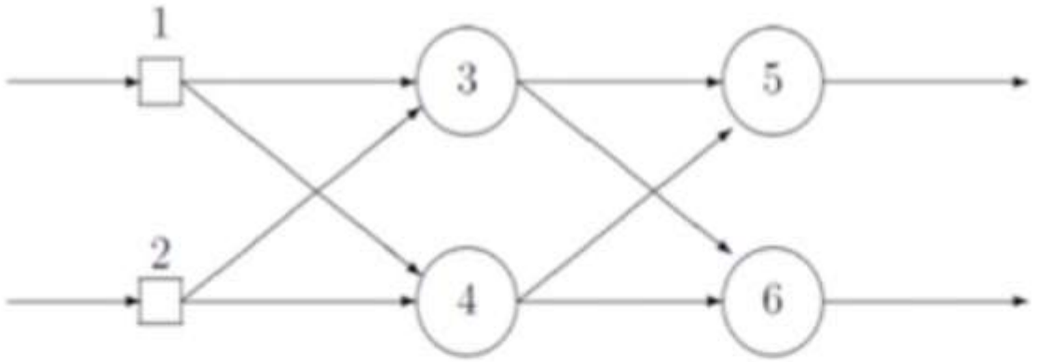


Consider the above Bayesian Belief Network,

a). Suppose we observe the fact that the grass is wet. There are two possible causes for this: either it rained, or the sprinkler was on. Which one is more likely?

b). It is cloudy, what's the probability that the grass is wet?

The following diagram represents a feed-forward neural network with one hidden layer:



A weight on connection between nodes i and j is denoted by w_{ij}, such as w_{13} is the weight on the connection between nodes 1 and 3. The following table lists all the weights in the network:

$w_{13} = -2$	$w_{35} = 1$
$w_{23} = 3$	$w_{45} = -1$
$w_{14} = 4$	$w_{36} = -1$
$w_{24} = -1$	$w_{46} = 1$

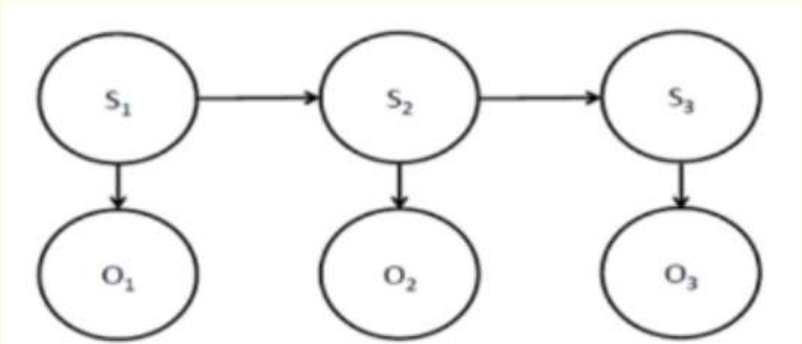
Each of the nodes 3, 4, 5 and 6 uses the following activation function:

$$\varphi(v) = \begin{cases} 1 & \text{if } v \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

where v denotes the weighted sum of a node. Each of the input nodes (1 and 2) can only receive binary values (either 0 or 1). Calculate the output of the network (y_5 and y_6) for each of the input patterns:

Pattern:	P_1	P_2	P_3	P_4
Node 1:	0	1	0	1

Suppose that we have binary states (labeled A and B) and binary observations (labeled 0 and 1) and the initial, transition, and emission probabilities as in the given table. Using the Veterbi algorithm, compute the probability that we observe the sequence O1 = 0, O2 = 1, and O3 = 0. Show all steps in solving the problem.



State	$P(S_1)$
A	0.99
B	0.01

(a) Initial probs.

S_1	S_2	$P(S_2\|S_1)$
A	A	0.99
A	B	0.01
B	A	0.01
B	B	0.99

(b) Transition probs.

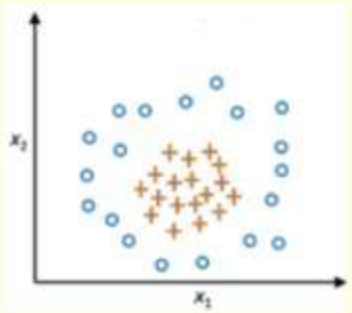
S	O	$P(O\|S)$
A	0	0.8
A	1	0.2
B	0	0.1
B	1	0.9

(c) Emission probs.

The following table contains dataset consists of factors that affecting sunburn Use ID3 algorithm to construct a minimal decision tree that predicts whether a person is affected by sunburn or not. **SHOW EACH STEP OF THE COMPUTATION**

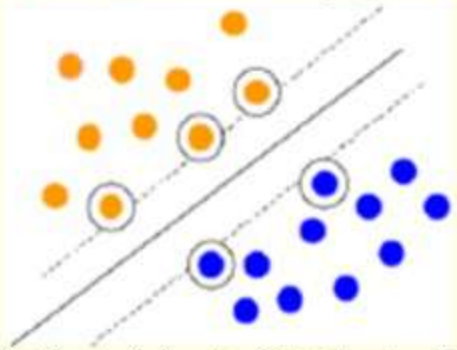
Name	*Hair*	*Height*	*Weight*	*Lotion*	*Result*
Sarah	blonde	average	light	no	sunburned (positive)
Dana	blonde	tall	average	yes	none (negative)
Alex	brown	short	average	yes	none(negative)
Annie	blonde	short	average	no	sunburned(positive)
Emily	red	average	heavy	no	sunburned(positive)
Pete	brown	tall	heavy	no	none(negative)
John	brown	average	heavy	no	none(negative)
Katie	blonde	short	light	yes	none(negative)

a) Let us design a model to predict the gender of a person: female or male based on the features like (length of hair, use of lipstick etc.) But, while plotting the data in a 2D plane, it seems as follows



where the data are randomly distributed. Explain how using Gaussian kernel, we can solve this non-linear problem with SVM. **(7 Marks)**

b) Consider the following figure to answer the following **(3 Marks)**



i) After training the SVM to classify male and female, can we discard all the other samples which are not support vectors and can still classify new examples? Justify your answer.
ii) Why SVM is called large margin classifier and what is the advantage of having large margin in a classifier.

a) Estimate the conditional probability of each attribute {Size, Color, Shape} to make a decision whether to buy an object or not. Using this probability, estimate the probability values for the new instance {Big, Pink, Circle} and thus, classify the instance. **(10 Marks)**

Sl. No.	Size	Color	Shape	Buy
1	Big	Yellow	Circle	Yes
2	Big	Yellow	Circle	No
3	Big	Yellow	Circle	Yes
4	Small	Yellow	Circle	No
5	Small	Yellow	Triangle	Yes
6	Small	Pink	Triangle	No
7	Small	Pink	Triangle	Yes
8	Small	Pink	Circle	No
9	Big	Pink	Triangle	No
10	Big	Yellow	Circle	Yes

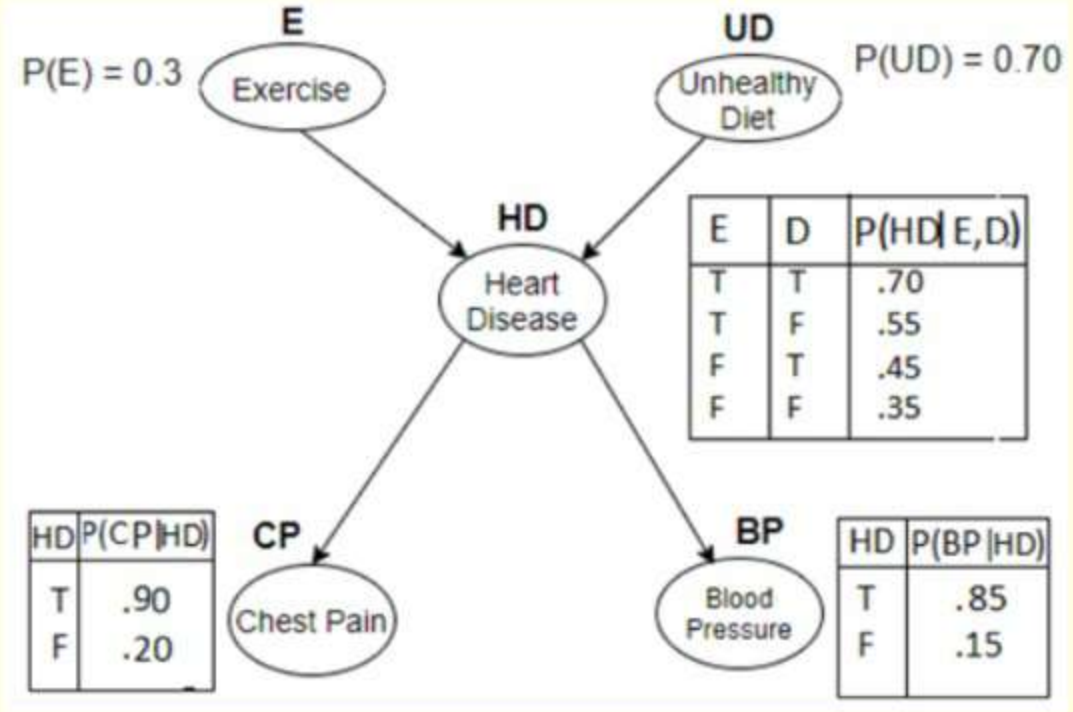
b) By considering the following Belief network, answer the following **(10 Marks)**

i) What is the probability that a person is having heart disease, chest pain ang high BP if he is not doing exercise and having unhealthy diet?

ii) What is the probability that a person is having high blood pressure, although he is not having Heart disease?

iii)What is the probability that a person is not having heart disease?

iv) What is the probability that a person is having chest pain, high blood pressure, heart disease without having unhealthy diet. But, the person is regularly doing exercise?



E	D	P(HD|E,D)
T	T	.70
T	F	.55
F	T	.45
F	F	.35

HD	P(CP|HD)
T	.90
F	.20

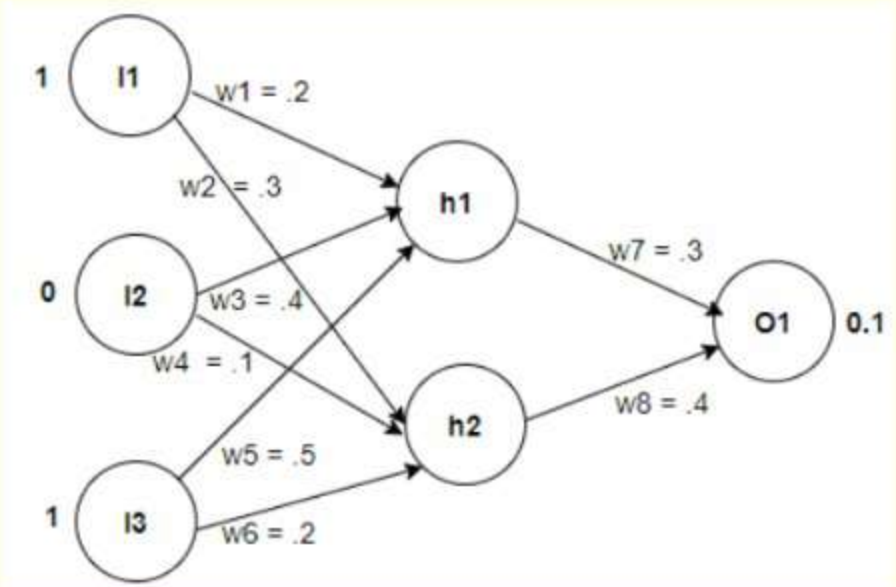
HD	P(BP|HD)
T	.85
F	.15

The following is a network where the output of each neuron in hidden and output layer is computed by applying the sigmoid function. Answer the following:

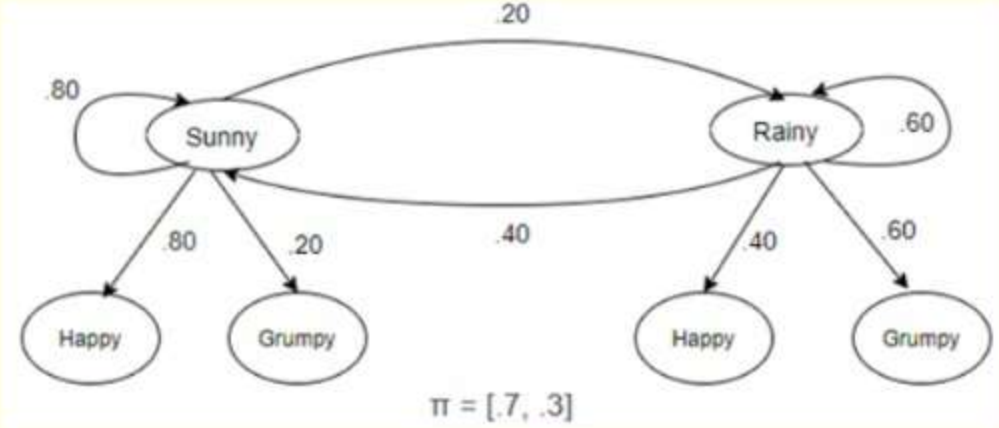
i. Compute the output of the hidden layer and output layer neurons for the given input (1, 0, 1).

ii. What is the error of this network with the given weights.

iii. Update the weights of w1, w7 and w8 by applying the backpropagation algorithm. Assume the Learning rate = 0.5.



Consider the state transition diagram which elaborates how a person feels on different climates. John is Happy Grumpy Happy from last 3 days. Looking at his moods by using Viterbi Algorithm, help his mother to predict the weather states in last three days.



What is Overfitting? Explain how regularization shrinks the parameter estimates towards zero in Linear Regression.

Build an ID3 Decision Tree to classify the following data.

#Sample	Model	Mileage	Color	Buy?
1	Old	High	Blue	No
2	Old	High	Blue	No
3	New	High	Blue	Yes
4	New	Moderate	Red	Yes
5	New	Moderate	Red	Yes
6	Old	Moderate	Blue	No
7	Old	High	Blue	No
8	Old	High	Red	Yes
9	New	Moderate	Red	Yes
10	New	Moderate	Blue	Yes

Suppose we have the following training set of positive (+) and negative (-) instances and one testing instance. Based on the provided descriptors: tumorsize and volume, we need to predict the malignancy of a tumor with given tumorsize and volume. Data including tumorsize, volume and type of tumor is shown below –

Instance	Tumor Size	Volume	Tumor Type (Malignant/ Benign)
1	2.00	2.00	B
2	4.25	3.75	B
3	4.00	4.00	B
4	3.00	3.25	B
5	4.50	5.00	B
6	7.25	5.75	M
7	7.50	8.00	M
8	5.50	6.75	M
9	4.75	6.25	M
10	7.00	4.25	M

Use K-NN algorithm and find the type of tumor with tumor size = 6.75 and volume = 3.00 with K= 5 and K = 7.

A second-hand car dealer has 10 cars for sale. The regression equation for these 10 cars with x = age of cars (years) and y = the mileage of cars (thousand miles) is as follows

$\hat{y} = 5.29+8.34x$

a) One car is 10 years old and its mileage is 85 thousand miles. What is the estimated mileage for this car based on the regression equation? What is the residual?

b) One car is 15 years old and its mileage is 128 thousand miles. What is the estimated mileage for this car based on the regression equation? What is the residual?

c) Find the mileage of cars with age 7 years and 12 years?

What is the size of the set of instances and hypothesis space for the following example? Apply Candidate Elimination algorithm on the given dataset below, find the learned version space.

Length: *Long, Short*
Color*: Green, Black*
Humidity*: Ink, Ball*

Length	Color	Type	Buy
Long	Green	Ink	No
Short	Green	Ball	No
Short	Green	Ink	Yes
Long	Black	Ink	No
Short	Black	Ink	Yes

Explain how distance-weighted k-Nearest Neighborhood reduces the limitations of simple K-NN.