Machine Learning

HOMEWORK-1

1. Points : $\{(-2,-1), (1,1), (3,2)\}$

a) Least square regression line:

Let us begin to find the best m (slope) and b (y-intercept) that suits the data for

$$y = mx + b$$

For each (x,y), let us calculate x^2 and xy and get the mean values respectively

x	y	x^2	xy
-2	-1	4	2
1	1	1	1
3	2	9	6
$\Sigma x = 2$	$\Sigma y = 2$	$\Sigma x^2 = 14$	$\Sigma xy = 9$

Now, $m = \dfrac{N\Sigma xy - \Sigma x \Sigma y}{N\Sigma x^2 - (\Sigma x)^2}$

$$= \frac{3(9) - (2)(2)}{3(14) - (2)^2} = \frac{27-4}{42-4} = \frac{23}{38} = \boxed{0.605}$$

$$b = \frac{\Sigma y - m\Sigma x}{N}$$

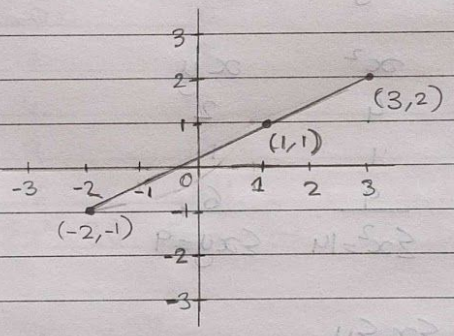
$$= \frac{2 - (0.605)(2)}{3} = \frac{2 - 1.21}{3} = \boxed{0.263}$$

Assemble the equation of the line, $y = mx + b$

$$y = 0.263x + 0.605$$

Now, error for the line can be calculated as,

x	y	$y = 0.263x + 0.605$	error
-2	-1	0.079	-1.079
1	1	0.868	0.132
3	2	1.295	0.705



2.

x	0	1	2	3	4
y	2	3	5	4	6

(a.) For the least square regression line $y = ax + b$, let's find the best solution for slope a and intercept b.

For each x & y, let us calculate x^2 and xy, then their respective means.

x	y	x^2	xy
0	2	0	0
1	3	1	3
2	5	4	10
3	4	9	12
4	6	16	24
$\Sigma x = 10$	$\Sigma y = 20$	$\Sigma x^2 = 30$	$\Sigma xy = 49$

Now, $$a = \frac{N\Sigma(xy) - \Sigma x \Sigma y}{N\Sigma(x^2) - (\Sigma x)^2}$$

$$= \frac{5(49) - (10)(20)}{5(30) - 10^2}$$

$$= \frac{245 - 200}{150 - 100} = \frac{45}{50} = \boxed{0.9}$$

$$b = \frac{\Sigma y - m\Sigma x}{N}$$

$$= \frac{20 - (0.9)(10)}{5} = \frac{11}{5} = \boxed{2.2}$$

Assemble the equation of the line, $y = ax + b$

$$\boxed{y = 0.9x + 2.2}$$

Now, error for the line can be calculated as,

x	y	$y = 0.9x + 2.2$	error
0	2	2.2	-0.2
1	3	3.1	-0.1
2	5	4.0	1
3	4	4.9	-0.9
4	6	5.8	0.2

(b) $x = 10$

$y = (0.9)(10) + 2.2$

$y = 11.2$

3. According to the Enjoy Sports learning task,

	Attributes	Values	Count
1.	Sky	Sunny, Cloudy, Rainy	3
2.	Temp	Warm, Cold	2
3.	Humid	Normal, High	2
4.	Wind	Strong, Weak	2
5.	Water	Warm, Cool	2
6.	Forecast	Same, Change	2

So, the distinct observations in $X = 3 \times 2 \times 2 \times 2 \times 2 \times 2 = 96$ and then, for hypothesis representation, value of each attribute can either be "?" or "ϕ" other than the defined values.
Hence,

Distinct hypothesis in $H = 5(4^5) = 5120$

These 5120 combinations are syntactically different but when the attributes have null (ϕ) / empty values the combinations will be same semantically.

So, semantically distinct hypothesis in $H = 4(3)^5 = 972$.
Additionally, we consider an empty/null set, so $H = 972 + 1 = 973$.
When we add another attribute "Water current" which can take three values (Light, Moderate, Strong), the possible observations $= 96 \times 3 = 288$ and hypothesis in $H = 972 \times 4 = 3888$. and an empty set, thus $H = 3889$.

In general,

number of possible instances = 96K

where K is the possible values for additional attribute A.

Hypothesis in H = 972(K+1)+1

4(a) Candidate-Elimination Algorithm

$S_0 = \{ \langle \phi, \phi, \phi, \phi \rangle, \langle \phi, \phi, \phi, \phi \rangle \}$

$S_1 = \{ \langle$ male, brown, tall, US$\rangle \langle$ female, black, short, US$\rangle \}$

$S_2 = \{ \langle$ male, brown, ?, ?$\rangle \langle$ female, black, short, US$\rangle \}$

$S_3 = S_2 = \{ \langle$ male, brown, ?, ?$\rangle \langle$ female, black, short, US$\rangle \}$

$S_4 = \{ \langle$ male, brown, ?, ?$\rangle \langle$ female, ?, short, ?$\rangle \}$

$G_4 = \{ \langle$ male, ?, ?, ?$\rangle \langle ?, ?, ?, ? \rangle \}$

$G_3 = \{ \langle$ male, ?, ?, ?$\rangle \langle ?, ?, ?, ? \rangle \}, \{ \langle ?, ?, ?, ? \rangle \langle ?, ?, ?,$ US$\rangle \}$

$G_2 = \{ \langle ?, ?, ?, ? \rangle \langle ?, ?, ?, ? \rangle \}$

$G_1 = \{ \langle ?, ?, ?, ? \rangle \langle ?, ?, ?, ? \rangle \}$

$G_0 = \{ \langle ?, ?, ?, ? \rangle \langle ?, ?, ?, ? \rangle \}$

(b.) {⟨male, black, short, Portugese⟩ ⟨female, blonde, tall, Indian⟩}

The above hypothesis space with the given training example (positive) will have the following distinct hypothesis

$2^8 = 256$!

For the positive training data, the eight attributes can either have the given value respectively or ?, thus 2 values for 8 attributes.

c. Assuming that the learner has encountered only the positive example from part (b), for the learner to converge to single correct hypothesis he will need a query for each attribute, meaning 8 queries for 8 attributes.

The queries are as follows:

{<female, black, short, Portuguese> <female, blonde, tall, Indian>}
{<male, brown, short, Portuguese> <female, blonde, tall, Indian>}
{<male, black, tall, Portuguese> <female, blonde, tall, Indian>}
{<male, black, short, Indian> <female, blonde, tall, Indian>}
{<male, black, short, Portuguese> <male, blonde, tall, Indian>}
{<male, black, short, Portuguese> <female, black, tall, Indian>}
{<male, black, short, Portuguese> <female, blonde, short, Indian>}
{<male, black, short, Portuguese> <female, blonde, tall, Portuguese>}

The number of valid hypotheses will reduce to half after each query. (where hypotheses are from part b)

d. Considering that there are n number of instances in our observation space X.
The hypothesis space will be tested by the following number of queries.

$2^{|X|-1}$.

5. For function: $X \rightarrow Y$

where $X = \langle x_1, x_2 \cdots x_n \rangle$ & x_i, Y are Boolean valued variable, let us consider a hypothesis space H where

if $[(x_i = a) \wedge (x_j = b)]$ then $Y = 1$ or 0.

Let $i = 1$ & $j = 2$,

x_1	x_2
0	0
0	1
1	0
1	1

for each choice of x_i & x_j we have 4 choices.

For each choice of i and j (2 variables) there are a total of n options, which is nC_2.

Distinct hypotheses in $H = \boxed{4\ nC_2}$.