Homework-1 Machine Learning CS6375.002

1. Set of pts given : $\{(-2,-1), (1,1), (3,2)\}$.
We find eq^{n}s of lines for every 2 sets of pts and calculate the LSE for the third pts.

⇒ Line connecting $(-2,-1)$ $(1,1)$.

$$y - y_1 = \left(\frac{y_2 - y_1}{x_2 - x_1}\right)(x - x_1)$$

$$(y-1) = \frac{-1-1}{-2-1}(x-1) \rightarrow y-1 = \frac{-2}{-3}(x-1)$$

$$3y - 3 = 2x - 2$$

$$\boxed{2x - 3y + 1 = 0}$$

~~Line connecting~~. LSE for $(3,2)$ for this line :.
We put $x = 2$ in ~~given~~ calculated eq^n of line & get corresponding value of y:

$$y = \frac{2x+1}{3} \quad \text{for } x = 2,\ y' = \frac{5}{3}.$$

$$LSE = \left(3 - \frac{5}{3}\right)^2 = \left(\frac{4}{3}\right)^2 = \frac{16}{9} = \boxed{1.778}$$

⇒ Line connecting $(1,1)$ & $(3,2)$.

$$y - 2 = \frac{-1}{-2}(x-3).$$

$$x - 3 = 2y - 4.$$

$$\boxed{x - 2y + 1 = 0}$$

LSE for pt $(-2,-1)$.

$$y' = \frac{x+1}{2} = \frac{-2+1}{2} = \frac{-1}{2}.$$

$$LSE = \left(-1 - \left(\frac{-1}{2}\right)\right)^2 = \left(\frac{-1}{2}\right)^2 = \frac{1}{4} = \boxed{0.25}$$

Line connecting $(-2,-1)$ & $(3,2)$.

$$y-2 = \frac{-1-2}{-2-3}(x-3)$$

$$y-2 = \frac{-3}{-5}(x-3)$$

$$5y-10 = 3x-9$$

$$\boxed{3x-5y+1=0}$$

LSE for $(1,1)$

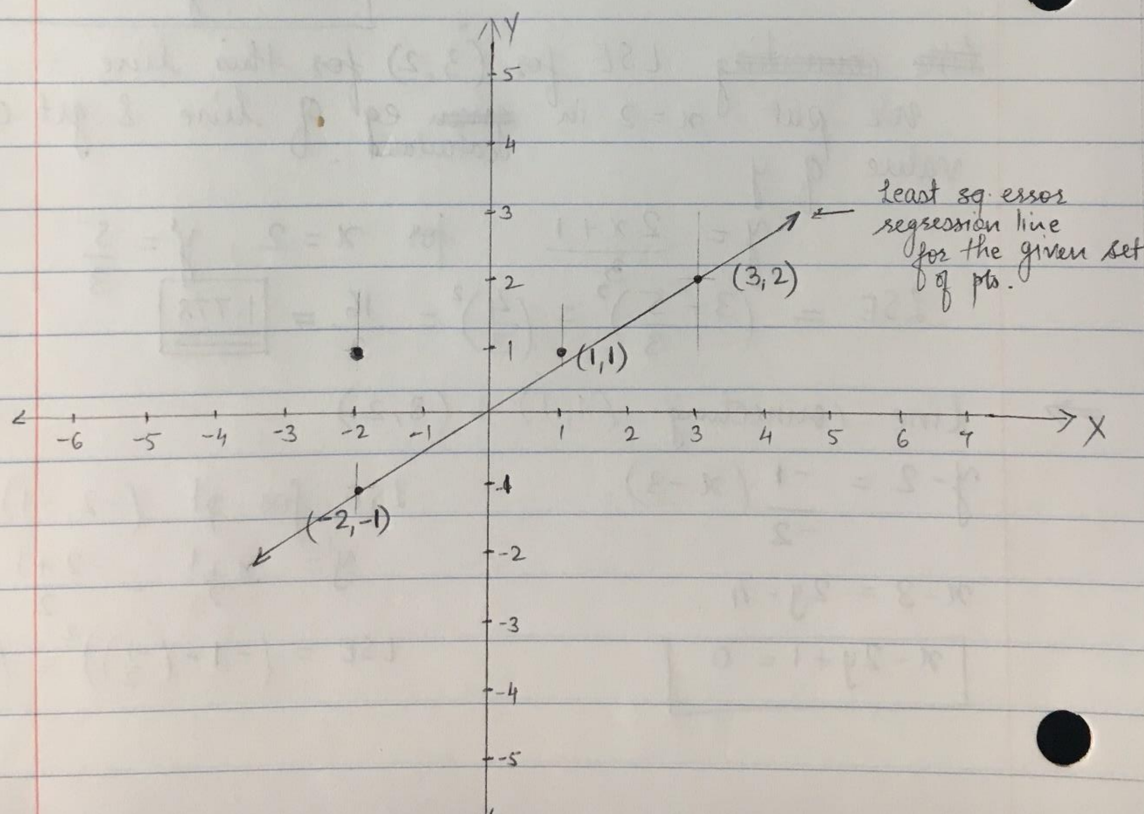
$$y' = \frac{3x+1}{5}$$

$$y' = \frac{4}{5}$$

$$LSE = \left(1-\frac{4}{5}\right)^2 = \frac{1}{25} = 0.04$$

The least sq. error is given by the line joining $(-2,-1)$ & $(3,2)$ so that becomes our least square regression line for the given data pts.

$$\boxed{3x-5y+1=0}$$



2.

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

We assume line $\to$ $W_0 + W_1 x$ to the least sq. regression line for the given data pts

$$J(W_0, W_1) = \sum_{i=1}^{5} \left(y_i - (W_0 + W_1 x_i) \right)^2$$

$$\frac{\partial J(W_0, W_1)}{\partial W_0} = 0 \quad \& \quad \frac{\partial}{\partial W_1} J(W_0, W_1) = 0$$

$$W_1 = \frac{m \sum_{i=1}^{m} x_i y_i - \sum_{i=1}^{m} x_i \sum_{i=1}^{m} y_i}{m \sum_{i=1}^{m} x_i^2 - \left(\sum_{i=1}^{m} x_i \right)^2}$$

$m \to$ no. of data pts.

$$W_0 = \frac{\sum_{i=1}^{m} y_i - W_1 \sum_{i=1}^{m} x_i}{m}$$

For the given data set.

$$\sum_{i=1}^{5} x_i y_i \Rightarrow 0 \times 2 + 1 \times 3 + 2 \times 5 + 3 \times 4 + 4 \times 6 = 49$$

$$\sum_{i=1}^{5} x_i = 0 + 1 + 2 + 3 + 4 = 10.$$

$$\sum_{i=1} y_i = 2 + 3 + 5 + 4 + 6 = 20.$$

$$\sum_{i=1}^{m} x_i^2 = 0^2 + 1^2 + 2^2 + 3^2 + 4^2 = 1 + 4 + 9 + 16 = 30$$

$m = 5$.

$$W_1 = \frac{5 \times 49 - (10 \times 20)}{5 \times 30 - (10)^2} = \frac{245 - 200}{150 - 100} = \frac{45}{50} = 0.9$$

$$W_0 = \frac{20 - 0.9 \times 10}{5} = \frac{11}{5} = 2.2$$

Eq^n of line $\Rightarrow$ $\boxed{2.2 + 0.9x = y}$ is the required line.

y at $x = 10 \Rightarrow \boxed{11.2}$

3. $\hat{y} = w_0 + w_1 x$ gives the least square error regression line for .

$$w_1 = \frac{m \overset{①}{\sum_{i=1}^{m} x_i y_i} - \overset{②}{\sum_{i=1}^{m} x_i} \overset{③}{\sum_{i=1}^{m} y_i}}{m \underset{④}{\sum_{i=1}^{m} x_i^2} - \left(\sum_{i=1}^{m} x_i\right)^2}$$

$$w_0 = \frac{\sum_{i=1}^{m} y_i - w_1 \sum_{i=1}^{m} x_i}{m}$$

With this, if we have stats for the data :-

$\bar{x} \rightarrow$ mean of x. $\quad \bar{x} \times m \rightarrow$ ②

$\bar{y} \rightarrow$ mean of y. $\quad \bar{y} \times m \rightarrow$ ③ .

$$C_{xx} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 . \quad C_{yy} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \bar{y})^2 .$$

$$C_{xy} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) .$$

a] Minimal set of statistics to estimate w_1 ?

w_1 has co-variance of xy in its numerator & variance of x in its denominator

Thus $w_1 = \frac{C_{xy}}{C_{xx}}$ are sufficient statistics for computing w_1 .

b) $y = w_0 + w_1 x$.

$\sum$ over all the pts :-

$$\sum_{i=1}^{m} y_i = \sum_{i=1}^{m} w_0 + \sum_{i=1}^{m} w_1 x_i . \rightarrow$$ dividing by 'm' $\rightarrow$ no of data pts.

$$\frac{1}{m} \sum_{i=1}^{m} y_i = \frac{1}{m} [m \cdot w_0] + w_1 \frac{1}{m} \sum_{i=1}^{m} x_i .$$

$$\boxed{\bar{y} = w_0 + w_1 \bar{x}}$$

$w_0 = \bar{y} - w_1 \bar{x}$. $\quad w_1$ needs C_{xy} & C_{xx} .

In addition we need $\bar{x}$ & $\bar{y}$.

Thus, we need $C_{xy}, C_{xx}, \bar{x}$ & $\bar{y}$.

c. Adding a new data pt x_{n+1}, y_{n+1}.

$$\bar{x}^{(n+1)} = \frac{1}{n+1} \sum_{i=1}^{n+1} x_i \quad \text{— def}^n \text{ of mean.}$$

$$= \frac{1}{n+1}\left[\sum_{i=1}^{n} x_i + x_{n+1}\right].$$

$$\bar{x}^{(n+1)} = \frac{1}{n+1}\left[n\bar{x}^{(n)} + x_{n+1}\right].$$

$$= \frac{1}{n+1}\left[n\bar{x}^{(n)} + \bar{x}^{(n)} + x_{n+1} - \bar{x}^{(n)}\right]$$

$$\bar{x}^{(n+1)} = \bar{x}^{(n)} + \frac{1}{n+1}\left[x_{n+1} - \bar{x}^{(n)}\right].$$

New mean = Old mean + $\frac{1}{\text{new data pts}}$ (new x − old mean)

$$\bar{y}^{(n+1)} = \frac{1}{n+1}\left(\sum_{i=1}^{n+1} y_i\right).$$

$$\bar{y}^{(n+1)} = \frac{1}{n+1}\left(\sum_{i=1}^{n} y_i + y_{n+1}\right).$$

$$= \frac{1}{n+1}\left(n\bar{y}^{(n)} + \bar{y}^{(n)} + y_{n+1} - \bar{y}^{(n)}\right).$$

$$= \bar{y}^{(n)} + \frac{1}{n+1}\left(y_{n+1} - \bar{y}^{(n)}\right)$$

$$\bar{y}^{(n+1)} = \bar{y}^{(n)} + \frac{1}{n+1}\left(y_{n+1} - \bar{y}^{(n)}\right).$$ Similar update to $\bar{x}^{(n+1)}$!

4. EnjoySport has 6 attributes which can take possibly values as :-

i)

Sky	Temp.	Humid.	Wind	Water	Forecast
Sunny	Warm	Normal	Strong	Warm	Same
Rainy.	Cold.	High.	Weak	Cool.	Change.
Cloudy					
3	2	2	2	2	2.
4	3	3	3	3	3.

+ ?

Hypothesis space cardinality is the number of different combination of the possible values for each attribute represented in the hypothesis space.
The values in the hypothesis space for each attribute include the specific values that can be assigned in addition to the '?' & ϕ.

So Total combinations $\rightarrow 4 \times 3 \times 3 \times 3 \times 3 \times 3 = 972$.
+ 1 $\longrightarrow$ for the 'ϕ' symbol.

(Since the 'ϕ' symbol for any attribute always classifies an instance as -ve, we count all the hypotheses with this value only once).
Thus, the space is $972 + 1 = 973$.

ii) Adding water current attribute with 3 possible values would mean all possible combⁿ $\Rightarrow$. $4^2 \times 3^5 = 3888$.
+ 1

Possible Hypotheses $\leftarrow$ $\underline{3889}$.

possible instances = $3^2 \times 2^5 = 288$ instances.

iii) Adding a new attribute with 'K' possible values : changes.

① No. of instances → $\boxed{K \times (\text{original instances})}$

② No. of Hypotheses → $(K+1)(\text{Original Hypothesis} - 1) + 1$.

$\Rightarrow \boxed{(K+1) \times (\text{Original no}) - K}$.

Check → K=3 original = 973 ∴ $4 \times 973 - 3 = 3889$ ✓

5- Attributes for each person :-

Sex.	Hair Color.	Height	Nationality
Male	Black	Tall	US, French - - ·
Female	Brown	Medium	- - . Portugese.
	Blonde.	Short.	
2.	3	3	7

		Y
Training Samples.	<Male, Brown, Tall, US> <Female Black Short US>	+
	<Male, Brown, Short, French> <Female Black Short US>	+
	<Female Brown Tall German> <Female Black Short Indian>	-
	<Male Brown Tall Irish> <Female Brown Short Irish>	+

Initial State.

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

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

Training eg. 1

S_1 {<Male Brown Tall US> <Female Black Short US>}

G_1 {< ? ? ? ?> <? ? ? ?>}

Training Example 2

S_2 {<Male, Brown ? ?> <Female Black Short US>}

G_2. {< ~~Male~~ ? ~~Brow~~ ? ? ?> < ? ? ? ? >}

Training Example 3

S_3 {< Male Brown ? ?> <Female Black Short US >}

G_2 {<<Male ? ? ?> <? ? ? ? >>
<< ? ? ? ?> <? ? ? US>>}

Training Example 4.

S_4 {<Male Brown ? ?> <Female ? Short ? >}

G_4. {<<Male ? ? ?> <? ? ? ?>>}

b). With the positive training example.
<<Male Black Short Portuguese> <Female Blonde Tall Indian>

Each hypothesis consistent with the above example can have either the specified value seen above or "?" for each attribute. That would mean for each of the 8 attributes a possibility of 2 options

So $2^8 = 256$

c. S_1 :- < Male Black Short Portugeese> < Female Blonde Tall Indian >.

Query 2 :

< Female Brown Mediam US> < Male Black Medium US>

→

S_2 << (? or male) (? or black) (? or short) (? or Portuguese)>
< ? or female> < ? or blonde> < ? or tall >< ? or Indian>>

Here Sex Attribute of the Hypothesis is determined. There are no more values it could assume to force S to generalize that attribute, if it hasn't already.

Query 3 :

<< done Blonde Tall French> < done Brow Short French>>

S_3 : << done (? or black) (? or short) (? or Portuguese)>
< done (? or blonde) (? or tall) (? or Indian)>>

There are no more values for Hair Color or Height to force S to generalize those attributes.

Query 4 :

<< done done done Japanese >> < done done done Japanese>>

This goes on till are the Nationalities for each person have been tried (3 more queries). This makes a total of 7 queries.

2\. For each query, the hypothesis space was reduced by half. That is except for the second query where both the hair color attribute and the height attribute were guaranteed convergence. This is consistent with 2⁸ calculation for the number of total hypotheses consistent with the original training example.

d) Instead of checking just 2 possible values for attributes in the hypotheses, we would have to check for every combination of values over all the possible values in the instance space.

~~This would mean~~

$\cancel{3 \times 4 \times 4 \times 8} = \cancel{384}$ ~~Answer~~ ← ~~All possible values in the instance space~~

~~Sex~~ ~~Hair Color~~ ~~Height~~ ~~Nationality~~

To generate a query for each possible instance other than the one already seen:

That sequence will be of length

$$(2 \times 3 \times 3 \times 7) \times (2 \times 3 \times 3 \times 7) = \underline{15876}.$$

$X_1, X_2 \ldots X_N$.

Each hypothesis constrains exactly 2 features of X

Select 2 from N in $\boxed{{}^N C_2}$

& each attribute can take either 0 or 1.

$\rightarrow \quad 2 \times 2 = 4.$

4 hypotheses combinations for an attribute pair.

$\therefore$ Total Hypotheses $= \boxed{{}^N C_2 \times 4}$.

Distinct.

$$= \frac{N(N-1)(N-2)!}{(N-2)! \times 2!} \times 4$$

$$= \boxed{2N(N-1)}.$$