Learning Paradigm - Regression

L Jeganathan

School of Computing Sciences and Engineering

So far, we have learnt...

- Necessity of Machine Learning.
- Nature of Problems that can be handled by Machine Learning.
- What is learning (in general)? - Different perspectives of Learning.
- A general definition of a Machine Learning Model.

OUTLINE

Land Price-fixing Problem

To fix a price for a rectangular land, given a few descriptions about the land.

How to solve this problem, through 'Learning'?

Requirements

- 'Experience' has to be made ready.
- We learn only through experience!

How to make ready the 'Experience (E)'

- Identify a few factors, on which, the price of the land depends.
- Some factors may be : length, breadth, locality (Residential(R) or Commercial(C)), distance between the land and the nearest hospital(dh), distance between the land and the nearest school(ds) etc.
- One can have any number of factors but the factors have to be independent of each other.
- Collect the information of the above factors for each of the land in the locality (where our land is situated) and get to know the prices quoted by the respective owners.
- For how many lands in that locality, have we to collect the information?

VIT Vellore Institute of Technology

A Typical Experience(E)

S.no.	length (l) in ft.	breadth(b) in ft.	Locality(L) (R or C)	dh in ft.	ds in ft.	Price(pr) in Lakhs
1	10	20	R	10	30	2
2	15	30	R	05	05	4.5
3	30	10	C	07	10	3
4	10	10	R	02	15	1
5	32	12	R	07	10	4
6	20	12	C	07	10	2.5
.	.	.	.	.	.	.
.	.	.	.	.	.	.

Factors : length, breadth, Locality, dh, ds are called as the input variables, independent variables attributes, parameters, features etc.
Price of the land : pr is called the Output variable, Dependent variable.

'Experience' with all the numerals

We code R as 1 and C as 0

S.no.	length(l) in ft.	breadth (b) in ft.	Locality(L) (1 or 0)	dh in ft.	ds in ft.	Price(pr) in Lakhs
1	10	20	1	10	30	2
2	15	30	1	05	05	4.5
3	30	10	0	07	10	3
4	10	10	1	02	15	1
5	32	12	1	07	10	4
6	20	12	0	07	10	2.5
.	.	.	.	.	.	.
.	.	.	.	.	.	.

What is the Task?

Task :

Given the information about all the input attributes of a land, To compute the price of the land.

For eg. If the input attributes are : Length - 29 ft, Breadth 24 ft, Locality is 0, dh is 1 ft., ds is 2 ft., what is the Price of the land?.

Task!

Given the input (29, 24, 0, 1, 2), what is the price, pr?

Note:

Our E does not have the input (29, 24, 0, 1, 2) .

How to accomplish the task (T)?

Procedure:

- Propose an equation (called as a hypothesis) that involves all the input variables.
  Let $\hat{pr}(x) = w_1 l + w_2 b + w_3 L + w_4(dh) + w_5(ds) + w_0$, where
  $x = (l, b, L, dh, ds)$
  $\hat{pr}(x)$ is the Price, predicted by our hypothesis, for the input x.

Learning Process called Regression

Considering all the values (both input and the output) of the training data, we compute the values of the unknowns (Weights): $w_0, w_1, w_2, w_3, w_4, w_5$

Assume that, with the training data, we have learnt $w_0, w_1, w_2, w_3, w_4, w_5$.

Procedure - continued

- Let $w_0 = 0, w_1 = 0.2, w_2 = 0.1, w_3 = 0.01, w_4 = 0.001, w_5 = 0.002.$
- $\hat{pr}(x) = (0.2)I + (0.1)b + (0.01)L + (0.001)(dh) + (0.002)(ds)$, where $x = (i, b, L, dh, ds)$
- If we have to compute the price for the input (29, 24, 0, 1, 2), then $\hat{pr}(29, 24, 0, 1, 2) = (0.2)29 + (0.1)24 + (0.01)(0) + (0.001)(1) + (0.002)(2) = 8.21$

Thus, we have learnt to predict the price of the land based on the experience E. T is accomplished based on E, through a learning process (model) called ‘Regression’.

How far our ‘learning model’ is correct in predicting the price of the land?

Performance measure of the Learning

To compute the value of 'P':

- For the first item in E: $x^1 = (l^1, b^1, L^1, dh^1, ds^1) = (10, 20, 1, 10, 30)$ Calculate $\hat{pr}^1$. We denote $\hat{pr}(x^1)$ by $\hat{pr}^1$.
- Price predicted by our 'learning' for the first item, $\hat{pr}^1 = (0.2)(10) + (0.1)(20) + (0.01)(1) + (0.001)(10) + (0.002)(30) = 4.08$
- Price for the first item in 'E', pr^1 is 2.
- Hence, $\hat{pr}^1 - pr^1 = 4.08 - 2 = 2.08$.
- 2.08 is the difference between the predicted value (due to the learning) and the actual value in 'E'. This is called the error in the first instance.
- Sometimes this error may be positive or negative. Hence, we square this error.
- We calculate $(\hat{pr}^1 - pr^1)^2 = (4.08 - 2)^2 = (2.08)^2 = 4.37$
- Like this, we calculate the error in all the N instances of 'E'.

VIT
Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

Performance measure - continued

- We calculate the squared error for all the instances and take the average.
- We compute the performance in terms of the error made in the 'learning'.

Hence, Performance Measure of the Learning,

$$P = \frac{\sum_{t=1}^{N} (\hat{pr}^t - pr^t)^2}{N}$$

Note

'P=0' means that, our learning model had learnt perfectly and predicted exactly.

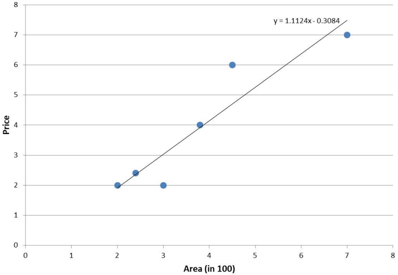
P is high $\implies$ Our learning model had not learnt nicely.
Then, we try to learn through another hypothesis, say

$$\hat{pr}(x) = w_1 l^2 + w_2 b^2 + w_3 L + w_4(dh) + w_5(ds) + w_0.$$

Geometrical Interpretation- Fit a curve for the E

Consider an E, for the Land Price-fixing problem, with Area of the land as the only input attribute.

S.No.	Area in ft. in (Thousands)	Price(pr) (in lakhs)
1	2	2
2	4.5	6
3	3	2
4	7	7
5	3.8	4
6	2.4	2.4



Experience (E):

The input data should be of the form

$$X = \{x^t, r^t\}_{t=1}^{N}, \quad where \quad x = (x_1, x_2, ...x_d) \in R^d.$$

x is a d-dimensional vector. $r \in R$. Here, the x_i's are the input attributes. x is the input variable.r is the output variable.

$x^t = ({x_1}^t, {x_2}^t, ...x_d^t)$ is the t^{th}-item in the input data. r^t is the t^{th} output in the data.

Task(T):

To find the relationship that involves $x_1, x_2, ...x_d$ and r, by proposing a hypothesis with weights $w_0, w_1, ...w_d$

We have to learn the weights in such a way that the error made in the learning is minimum.

Performance Measure:

Average of the square of the errors (difference between the actual output and the predicted output) made in each of the instance of E.

Experience (E), in a general sense

S.no.	x_1	x_2	x_3	.	.	r
1	${x_1}^1$	${x_2}^1$	${x_3}^1$	.	.	r^1
2	${x_1}^2$	${x_2}^1$	${x_3}^2$	.	.	r^2
3	${x_1}^3$	${x_2}^3$	${x_3}^3$	.	.	r^3
.	.	.	.	.	.	.
. N	. ${x_1}^N$	. ${x_2}^N$	. ${x_3}^N$	. .	. .	. r^N

SUMMARY

Thus, we have understood

- The Process of Learning (in a general sense), called Regression. - yet to learn the actual ‘learning’ (model which will compute the weights involved in the hypothesis)
- Technical description of the problems for which regression model can be applied

EXERCISE

1.Consider the E given in the table. E is trained with two different hypothesish_1 and h_2 as follows:

$h_1(x) = w_1 A + w_0$, w_1= 1.1, w_0=0.98

$h_2(x) = w_1 A^2 + w_0$, w_1= 0.87, w_0=1.0

S.No.	Area(A)in ft. in (Thousands)	Price(pr) (in lakhs)
1	2	2
2	4.5	6
3	3	2
4	7	7
5	3.8	4
6	2.4	2.4

Which hypothesis learns better?

- 2. Propose a problem with a clear description of *E* and *T* where regression based learning is feasible.
- 3. For a problem, E is described. Every instance in E has one output variable. All the output variables of all the instances in E are not a numerical number. Is it possible to apply regression to that problem?.