Multivariate Regression

Dr. Prashant Singh Rana
psrana.com | psrana@gmail.com

Univariate

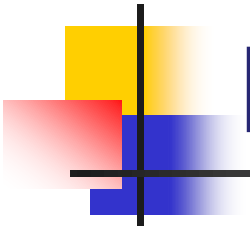
- One input and one output
- Example:
  - OTP per transaction: Every transaction have unique OTP

Transaction ID	OTP
3424234234	9456
5653453235	9879
5909087556	4536
8797890123	2345

Multivariate

- Multiple inputs and one output
- Example:
  - Cancer Prediction
  - Cement Mixture strength

x1	x2	x3	x4	x5	Strength
17	0	-5	0.784245	37	26
12	0	-10	0.587296	25	27
18	0	-7	0.876622	40	25
11	0	-7	0.80826	24	23
18	0	-4	0.83215	37	28
10	1	-9	0.62842	27	28
19	0	7	0.522811	44	30
19	-1	4	0.548609	37	23
15	0	-6	0.177904	46	20



Linear Model

Y = w1x1 + w2x2 + w3x3 + w4x4 + w5x5

Where,

- **w1, w2, w3, ... are the weights.**
- **x1, x2, x3, ... are the input features.**
- **We need to find the optimized weights.**

Linear Model

Input Data

C	F1	F2	F3	F4	F5
30	3	5	7	9	10
50	1	7	5	10	15
80	-15	15	10	5	10
.	.	.	.	.	.
.	.	.	.	.	.
90	20	0	2	3	7

Output

C = w1F1 + w2F2 + w3F3 + w4F4 + w5F5

Least Sum of Square

$$\text{Obj fun} = min\left(\sum_{i=1}^{n}\sqrt{\left(C_i - \sum_{j=1}^{D} w_j.F_{i,j}\right)^2}\right)$$

n = Total number of Observations (Rows)

C = Target (Class)

D = Dimension of the Problem (# of Features)

w = weight defined between [-1,1] or [-10,10]

F = Features (value of F1, F2, F3….)

Find Least Sum of Square

Problem 1

w	0.2	-0.3	0.1	-0.5	0.9

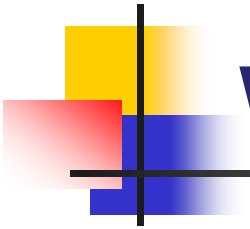
C	F1	F2	F3	F4	F5
30	3	5	7	9	10
50	1	7	5	10	15
80	-15	15	10	5	10
90	20	0	2	3	7

Find Least Sum of Square

Problem 2

w	-0.2	0.3	-0.1	0.5	-0.9

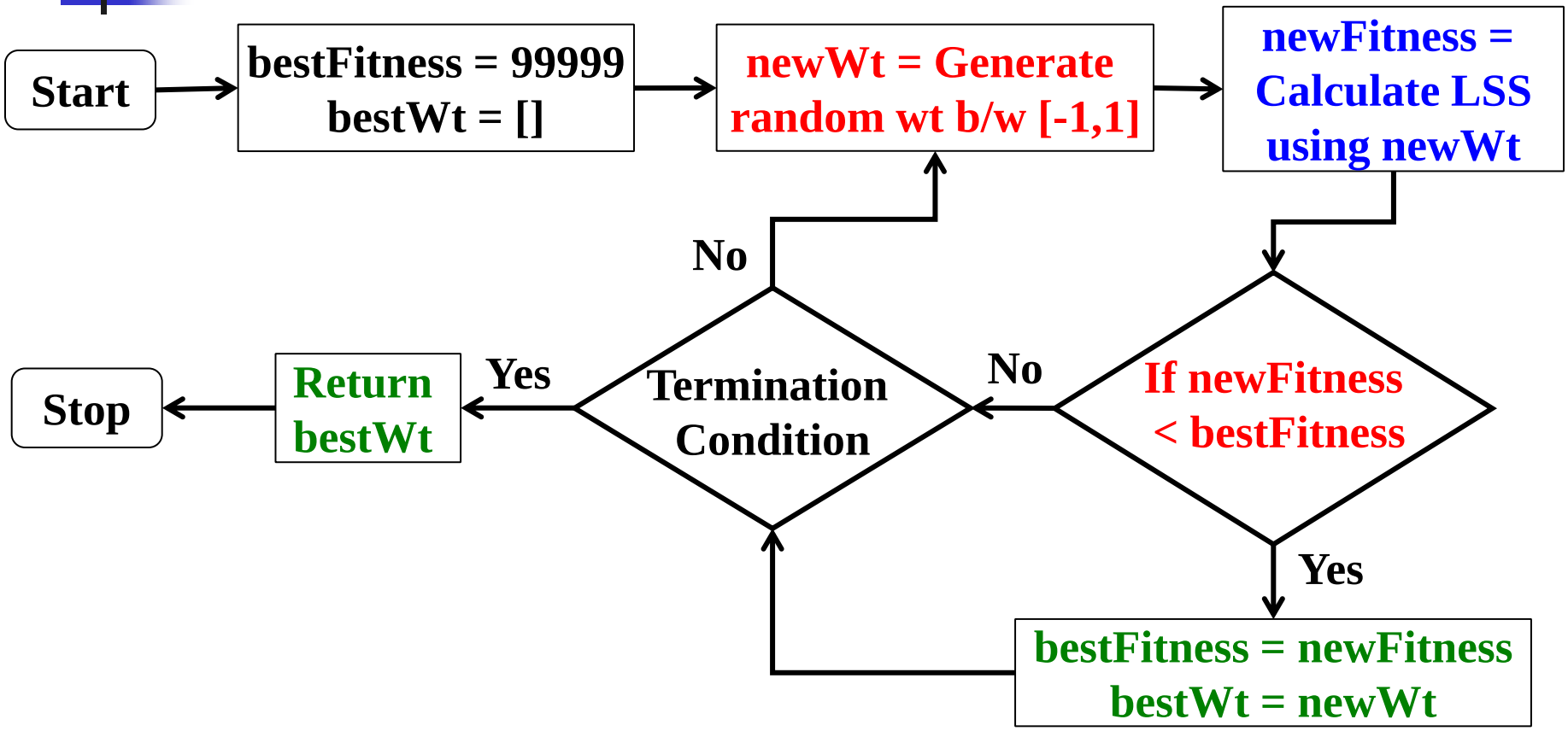
C	F1	F2	F3	F4	F5
30	3	5	7	9	10
50	1	7	5	10	15
80	-15	15	10	5	10
90	20	0	2	3	7



Which Wt. is better?

Problem 1 or Problem 2

Weight Optimization using Random Approach - min(LSS(wts))



Train/Test and Actual/Predicted

x1	x2	x3	x4	x5	Strength
17	0	-5	0.784245	37	26
12	0	-10	0.587296	25	27
18	0	-7	0.876622	40	25
11	0	-7	0.80826	24	23
18	0	-4	0.83215	37	28
10	1	-9	0.62842	27	28
19	0	7	0.522811	44	30
19	-1	4	0.548609	37	23
15	0	-6	0.177904	46	20

Train Data

Test Data

Train/Test and Actual/Predicted

x1	x2	x3	x4	x5	Strength
17	0	-5	0.784245	37	26
12	0	-10	0.587296	25	27
18	0	-7	0.876622	40	25
11	0	-7	0.80826	24	23
18	0	-4	0.83215	37	28
10	1	-9	0.62842	27	28
19	0	7	0.522811	44	30
19	-1	4	0.548609	37	23
15	0	-6	0.177904	46	20

Testing Data

input features

Actual

Put in the eq and get value (known as predicted)

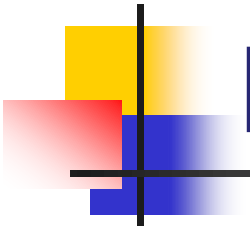
How to make predictions?

Approach:

1. **Find the weights using 70% Data.**
2. **Test the equation on 30% Data.**
3. **Evaluate the performance using actual and predicted on different parameters (such as correlation, RMSE, MAE, etc)**

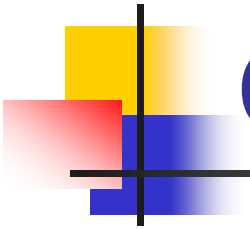
Develop Web application

- तुम मुझे Data दो, मैं तुम्हें Equation दूंगा
- Develop Web Interface



Next Class

- Implementation of LSS using Python.
- Implementation of Multivariate regression using scikit-learn library
- Handling of categorical values.



Question & Doubts