

BUILDING MACHINE LEARNING PROJECTS

7-STEP PROCESS

1) Form a hypothesis

1. Form a Hypothesis

We can predict how many medals a country will win in the Olympics.

2) Find the data

2. Find the Data

Team	Year	Athletes	Prev Medals	Medals
USA	2008	763	263	317
USA	2012	689	317	248
USA	2016	719	248	264
IND	2008	67	1	3
IND	2012	95	3	6
IND	2016	130	6	2

- 3) **Reshape the data-** identify and target column and predictors

3. Reshape the Data

Team	Year	Athletes	Prev Medals	Medals
USA	2008	763	263	317
USA	2012	689	317	248
USA	2016	719	248	264
IND	2008	67	1	3
IND	2012	95	3	6
IND	2016	130	6	2

- 4) **Clean the data-** remove the missing data

- 5) **Error metric-** Mean absolute error (actual value - predicted value)

5. Error Metric

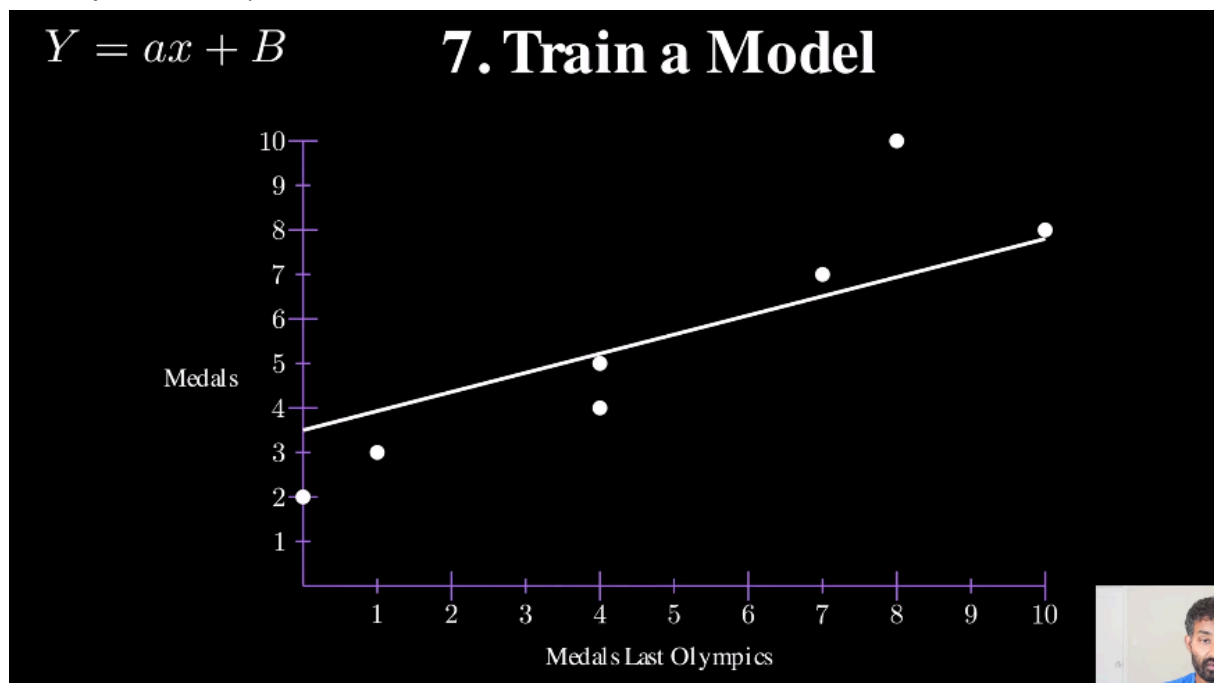
$$\sum_{i=1}^D |x_i - y_i|$$

Team	Year	Medals	Predictions	Error
ALB	1992	0	3	3
ALG	1964	0	2	2
AND	1976	0	2	2
BLR	1996	23	15	8
ARM	1996	2	5	3

6) **Split the data-** training data and test data

6. Split the Data				
Training Data	Team	Year	Athletes	Prev Medals
	USA	2008	763	263
	USA	2012	689	317
	IND	2008	67	1
Test Data	Team	Year	Athletes	Prev Medals
	USA	2016	719	248
	IND	2016	130	6

7) **Train the model-** using regression (linear regression draws a straight line to make predictions)



MULTIVARIATE WITH TWO PREDICTORS (Athletes and Medals last olympics)

Using Multivariate Regression,

$$Y = a_1x_1 + a_2x_2 + B$$