Hilal Işık Trial Lecture

Linear Regression

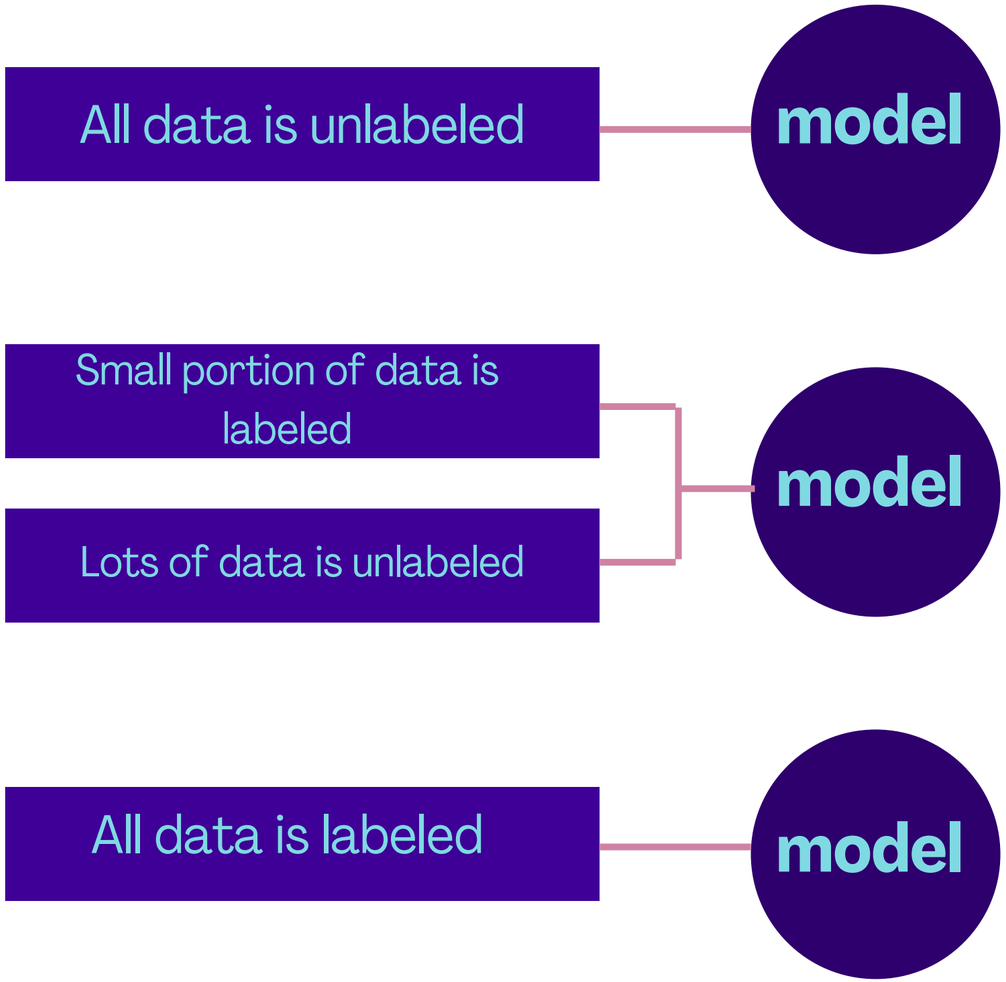
Understanding and Implementing

Lecture's Agenda

- What is Regression in Machine Learning?
- What is Linear Regression? Why and where to use it?
- Understanding Linear Regression
- How to implement Linear Regression using Python?

Types of Machine Learning

- Unsupervised
- Semi-Supervised
- Supervised



Regression

'investigates the relationship between a dependent and independent variable'

- Linear — Numerical Variable
- Logistic — Categorical variable

Linear Regression

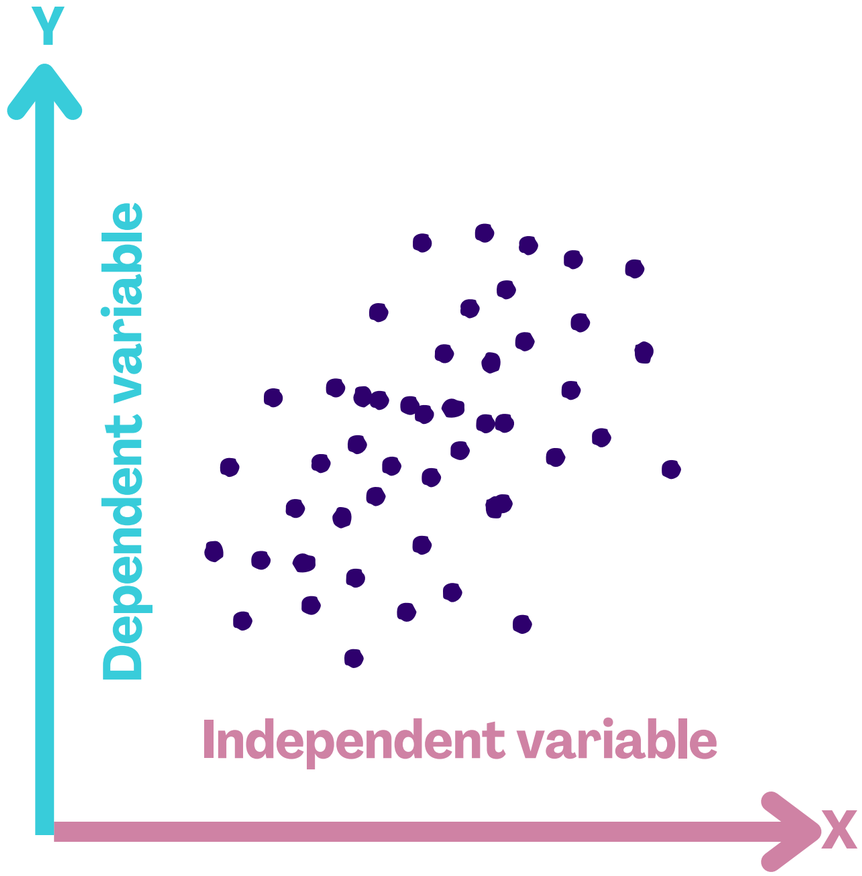
'to predict dependent variable (y) based on values of independent variables(X)'

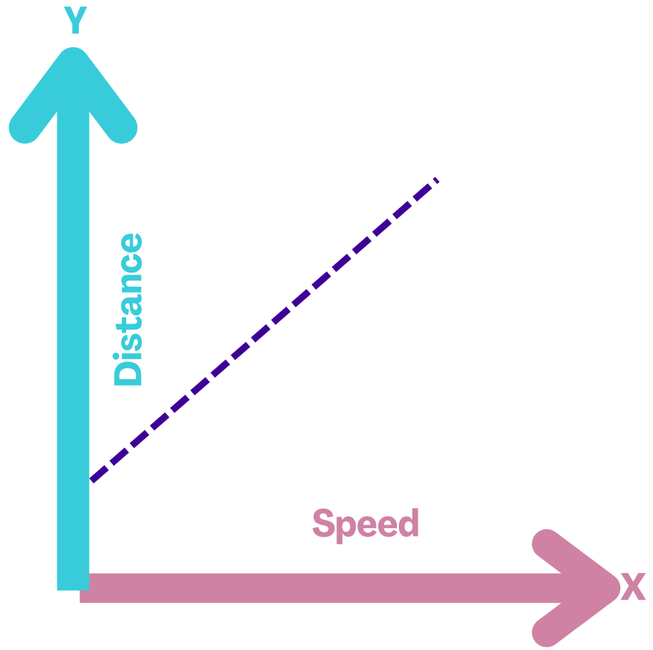
- computationally inexpensive
- easier to communicate

- Evaluating trends and sales estimates
- Analyzing the impact of price changes
- Assesment of risk

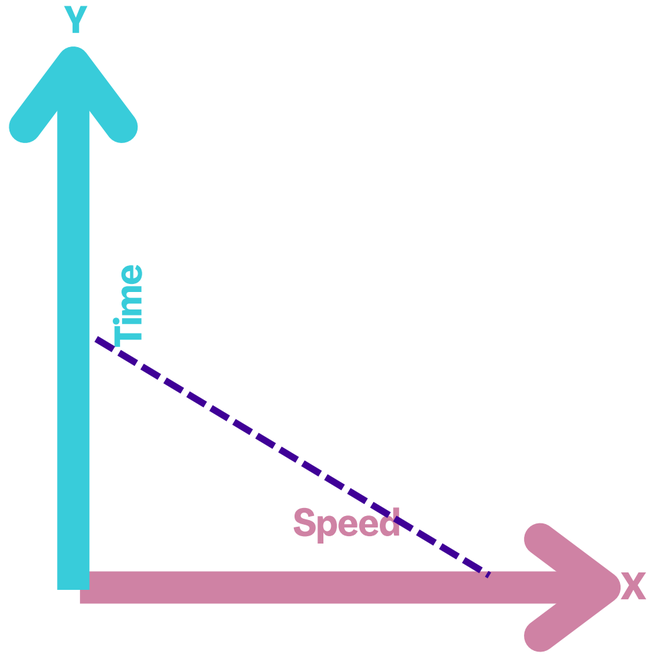
Linear Regression

$y = mx+c$





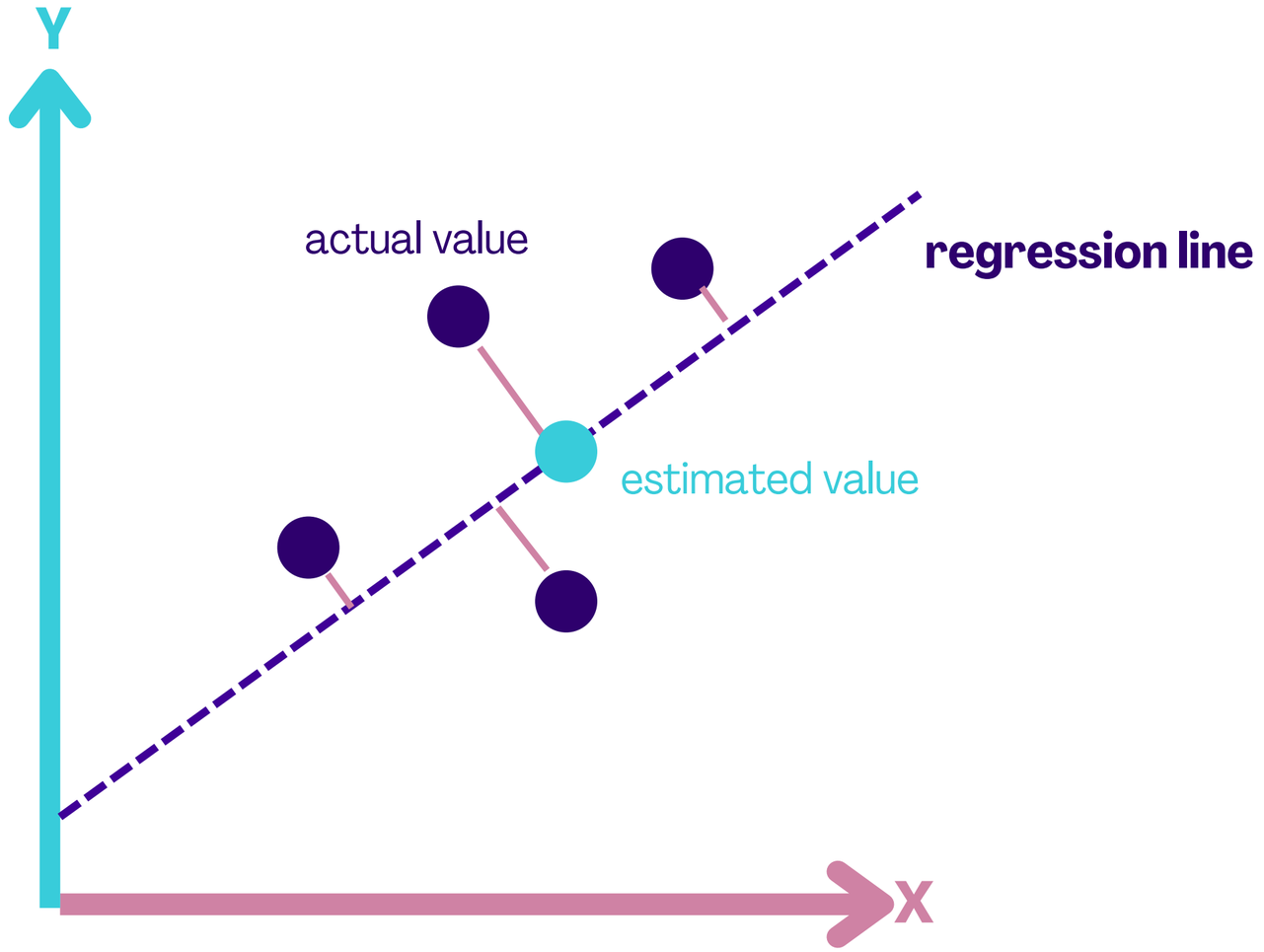
positive linear regression line



negative linear regression line

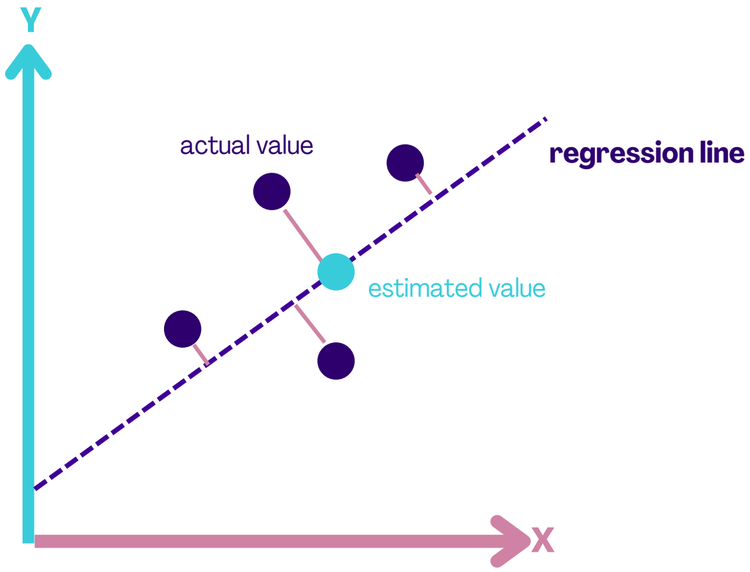
Linear Regression

$$y = mx+c$$



Linear Regression

$$MSE = \frac{1}{n}\sum\underbrace{\left(y - \hat{y}\right)^2}_{\text{The square of the difference between actual and predicted}}$$



$$MAE = \boxed{\frac{1}{n}}\sum\underbrace{\left|\boxed{y} - \boxed{\hat{y}}\right|}_{\text{The absolute value of the residual}}$$

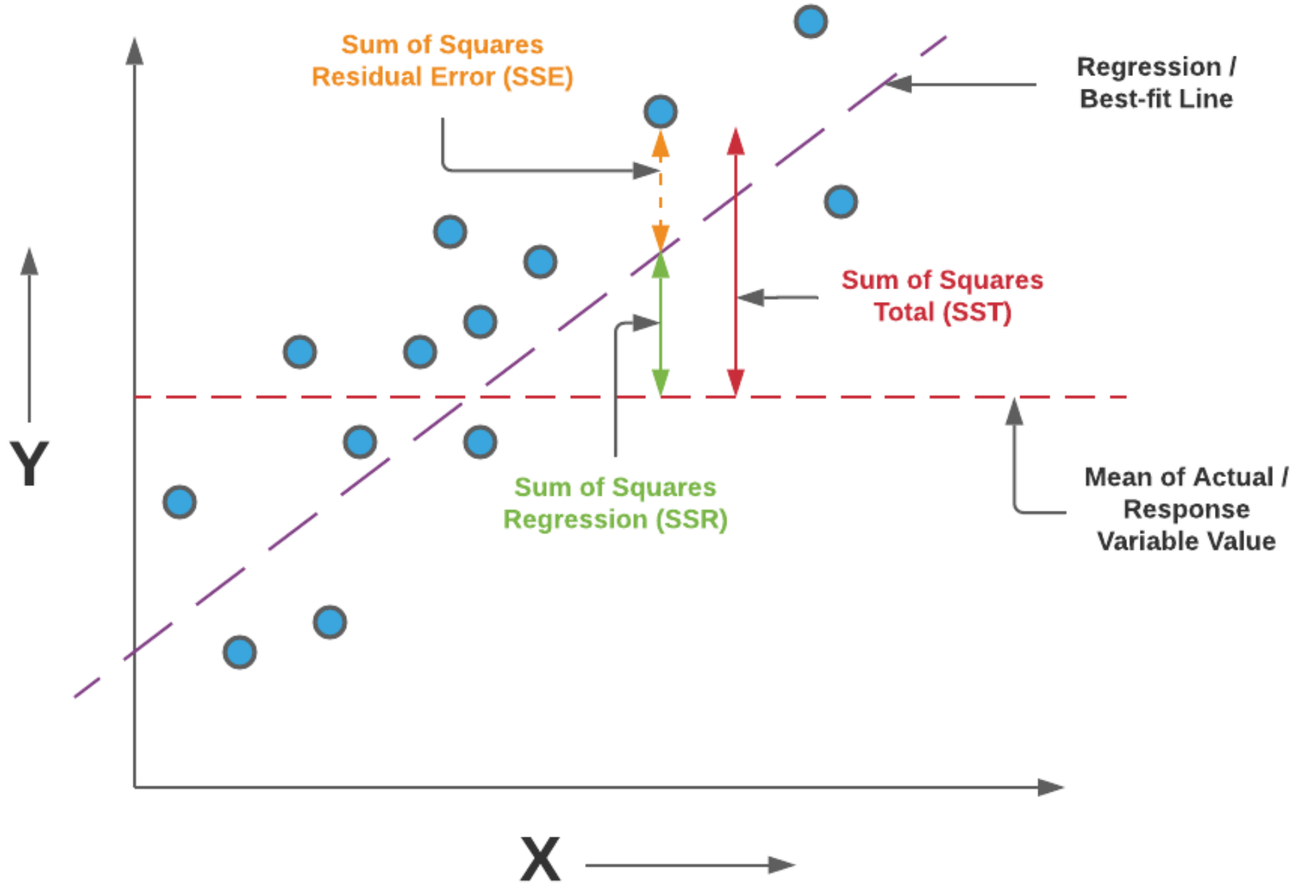
Divide by the total number of data points

Predicted output value

Actual output value

Sum of

Linear Regression



$$R^2 = 1 - \frac{SS_{RES}}{SS_{TOT}} = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \overline{y})^2}$$

Scikit-Learn

```
from sklearn.linear_model import LinearRegression
model = LinearRegression()
model.fit(X_train, y_train)
y_pred = model.predict(X_test)
```

Training and Test Splits

	Date	Title	Budget	DomesticTotalGross	Director	Rating	Runtime
0	2013-11-22	The Hunger Games: Catching Fire	130000000	424668047	Francis Lawrence	PG-13	146
1	2013-05-03	Iron Man 3	200000000	409013994	Shane Black	PG-13	129
2	2013-11-22	Frozen	150000000	400738009	Chris BuckJennifer Lee	PG	108
3	2013-07-03	Despicable Me 2	76000000	368061265	Pierre CoffinChris Renaud	PG	98
4	2013-06-14	Man of Steel	225000000	291045518	Zack Snyder	PG-13	143
5	2013-10-04	Gravity	100000000	274092705	Alfonso Cuaron	PG-13	91
6	2013-06-21	Monsters University	NaN	268492764	Dan Scanlon	G	107
7	2013-12-13	The Hobbit: The Desolation of Smaug	NaN	258366855	Peter Jackson	PG-13	161
8	2013-05-24	Fast & Furious 6	160000000	238679850	Justin Lin	PG-13	130
9	2013-03-08	Oz The Great and Powerful	215000000	234911825	Sam Raimi	PG	127
10	2013-05-16	Star Trek Into Darkness	190000000	228778661	J.J. Abrams	PG-13	123
11	2013-11-08	Thor: The Dark World	170000000	206362140	Alan Taylor	PG-13	120
12	2013-06-21	World War Z	190000000	202359711	Marc Forster	PG-13	116
13	2013-03-22	The Croods	135000000	187168425	Kirk De MiccoChris Sanders	PG	98
14	2013-06-28	The Heat	43000000	159582188	Paul Feig	R	117
15	2013-08-07	We're the Millers	37000000	150394119	Rawson Marshall Thurber	R	110
16	2013-12-13	American Hustle	40000000	150117807	David O. Russell	R	138
17	2013-05-10	The Great Gatsby	105000000	144840419	Baz Luhrmann	PG-13	143

Training Data

Test Data

IBM: Supervised Machine Learning: Regression

Training and Test Splits

```
# DEFINE X and y
y=df.col1
X=df.drop('col1', axis=1)

# SPLIT DATASET INTO TRAIN AND TEST
from sklearn.model_selection import
train_test_split
X_train, X_test, y_train, y_test
=train_test_split(X, y,test_size=0.3,
random_state=123)
```

Overfitting-Underfitting

overfitting simply means that the learning model is far too dependent on training data while underfitting means that the model has a poor relationship with the training data

Thank you so much!

hilal.hisik@gmail.com
https://github.com/isik-hilal