Regression

Profile

linkedin.com/in/romansyasetyo/

Table of Content

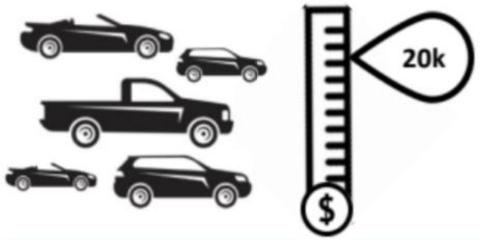
What will We Learn Today?

1. What is Regression
2. Examples of Regression Projects
3. Linear Regression
4. Decision Tree
5. Evaluation Metrics

Hands on using Python

REGRESSION

PREDICT VALUE



CLASSIFICATION

PREDICT CLASS

CLUSTERING

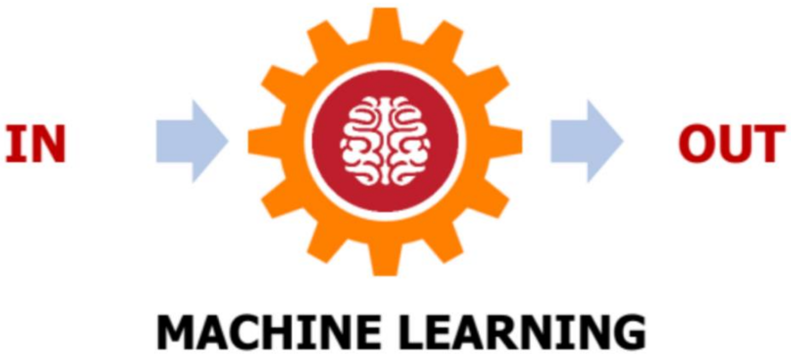
GROUP OBJECT

What is Regression?

Regression Model

Model the relationship between a dependent (target) and independent (predictor) variables with one or more independent variables

Value
Value
Value
.....



Value

Examples of Regression Projects

$82000

$55500

???

Advertisement	Sales
$90	$1000
$120	$1300
$150	$1800
$100	$1200
$130	$1380
$200	??

Regression : Linear Regression

Linear Regression

- Simple Linear Regression

  1 independent variable

  1 dependent variable

  $Y = \alpha + \beta x + \varepsilon$

- Multiple Linear Regression

  n independent variables

  1 dependent variable

  $Y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_n x_n + \varepsilon$

$\mathrm{Y} = \textit{predict value (variable independent)}$

$x = \textit{feature value (variable dependent)}$

$\alpha = \textit{intercept paramenter}$

$\beta = \textit{slope parameter}$

$\varepsilon = \textit{error (residual)}$

Simple Illustration

Dataset of Years of Experience

and Salary (in 1000$)

Independent Variables

Dependent Variables

Years of Experience	Salary in 1000$
2	15
3	28
5	42
13	64
8	50
16	90
11	58
1	8
9	54

Simple Illustration

Years of Experience x_i	Salary (in 1000$) y_i	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$
2	15	-5.56	-30.44	169.24	30.91
3	28	-4.56	-17.44	79.53	20.79
5	42	-2.56	-3.44	8.81	6.55
13	64	5.44	18.56	100.97	29.59
8	50	0.44	4.56	2.01	0.19
16	90	8.44	44.56	376.09	71.23
11	58	3.44	12.56	43.21	11.83
1	8	-6.56	-37.44	245.61	43.03
9	54	1.44	8.56	12.33	2.07
$\bar{x}$=7.56	$\bar{y}$= 45.44			∑ = 1037.8	∑ = 216.19

$$m = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$$

$$b = \bar{y} - m * \bar{x}$$

m = 1037.8 / 216.19

m = 4.80

b = 45.44 - 4.80 * 7.56 = 9.15

Hence, **y = mx + *b*** → ***4.80x + 9.15***

y = 4.80x + 9.15

Linear Regression in Python

```python
from sklearn.linear_model import LinearRegression

# Using Multiple Linear Regression
ols_model = LinearRegression().fit(X_train, y_train)

y_train_pred = ols_model.predict(X_train)
y_test_pred = ols_model.predict(X_test)
```

Linear Regression - Regularized

- Lasso Regression

```
# Using Lasso Regression
from sklearn.linear_model import Lasso

lasso_model = Lasso(alpha = 0.8, normalize = True).fit(X_train, y_train)

y_train_pred = lasso_model.predict(X_train)
y_test_pred = lasso_model.predict(X_test)
```

- Ridge Regression

```
# Using Ridge Regression
from sklearn.linear_model import Ridge

ridge_model = Ridge(alpha = 0.1, normalize = True).fit(X_train, y_train)

y_train_pred = ridge_model.predict(X_train)
y_test_pred = ridge_model.predict(X_test)
```

- Elastic-Net Regression

```
# Using Elastic Net Regression
from sklearn.linear_model import ElasticNet

elastic_net_model = ElasticNet(alpha=1, l1_ratio=0.5, normalize=False).fit(X_train, y_train)

y_train_pred = elastic_net_model.predict(X_train)
y_test_pred = elastic_net_model.predict(X_test)
```

Regression : Decision Tree

Tree-based Regressor

```
# Using Decision Tree Regressor
from sklearn.tree import DecisionTreeRegressor

dt_regressor = DecisionTreeRegressor().fit(X_train, y_train)

y_train_pred = dt_regressor.predict(X_train)
y_test_pred = dt_regressor.predict(X_test)
```

```
# Using Random Forest Regressor
from sklearn.ensemble import RandomForestRegressor

rf_regressor = RandomForestRegressor().fit(X_train, y_train)

y_train_pred = rf_regressor.predict(X_train)
y_test_pred = rf_regressor.predict(X_test)
```

Evaluation Metrics

Mean Absolute Error

- Regression metric which measures the **average magnitude of errors** in a group of predictions, without considering their directions.

$$\text{MAE} = \frac{1}{n}\sum_{t=1}^{n} |e_t|$$

Mean Squared Error

- **Average squared difference between the predictions and expected results**. In other words, an alteration of MAE where instead of taking the absolute value of differences, they are squared.

$$\mathrm{MSE} = \frac{1}{n}\sum_{t=1}^{n} e_t^2$$

DigitalSkola

Uncover The World Of Digital Skills With Us

Thank You