

Lab program no 5

Build Logistic Regression Model for a given dataset

Predicting if a person would buy life insurance based on his age using logistic regression

- `import pandas as pd`
- `from matplotlib import pyplot as plt`
- `%matplotlib inline`

- `df = pd.read_csv("insurance_data.csv")`
- `df.head()`

- `plt.scatter(df.age,df.bought_insurance,marker='+',color='red')`
- `from sklearn.model_selection import train_test_split`
- Perform the split
- `X_train, X_test, y_train, y_test = train_test_split(df[['age']],df.bought_insurance,train_size=0.8)`
- Display `X_test`
- `from sklearn.linear_model import LogisticRegression`
- `model = LogisticRegression()`
- `model.fit(X_train, y_train)`
- Display `X_test`

- `y_predicted = model.predict(X_test)`
- `model.predict_proba(X_test)`
- `model.score(X_test,y_test)`
- Display `y_predicted`
- Display `X_test`

Analyze thro Linear regression

- **model.coef_** indicates value of m in $y=m*x + b$ equation
- model.coef_
- **model.intercept_** indicates value of b in $y=m*x + b$ equation
- model.intercept_

Analyze thro sigmoid function

```
import math  
def sigmoid(x):  
    return 1 / (1 + math.exp(-x))
```

prediction_function

```
def prediction_function(age):  
    z = 0.042 * age - 1.53 # 0.04150133 ~ 0.042 and -1.52726963 ~ -1.53  
    y = sigmoid(z)  
    return y
```

Check for

age = 35

prediction_function(age)

Check the value with probability

age = 43

prediction_function(age)

Check the value with probability

- Download employee retention dataset from here: <https://www.kaggle.com/giripujar/hr-analytics>.
- 1. Now do some exploratory data analysis to figure out which variables have direct and clear impact on employee retention (i.e. whether they leave the company or continue to work)
- 2. Plot bar charts showing impact of employee salaries on retention
- 3. Plot bar charts showing correlation between department and employee retention
- 4. Now build logistic regression model using variables that were narrowed down in step 1
- 5. Measure the accuracy of the model

- https://github.com/codebasics/py/blob/master/ML/7_logistic_reg/Exercise/7_logistic_regression_exercise.ipynb