

Analysis and Prediction of Data Based on Linear Regression

D. A. Khalilov¹, M. J. Holmurzaeva²

Abstract: This article covers the essence of linear regression models, their practical applications and their role in data analysis. The paper built models based on Real-world data sets and estimated prediction accuracy and efficiency. The results of the study show that linear regression helps in the determination of continuous variables.

Keywords: regression, linear regression, machine learning, prediction,, R2, MSE Python, Machine Learning Data Analysis.

Introduction

In recent years, artificial intelligence and machine learning technologies have been widely used in a variety of fields, including economics, medicine, education, marketing, and industry. The role of regression models in the process of analyzing data and making predictions on their basis is incomparable. Linear regression is a statistical model used to predict continuous values to determine the linear relationship between input factors and output. Logistic regression, on the other hand, is used to determine results of two or more categories.

In this article:

- In Uzbekistan (California housing data collection),
- Issues of model evaluation methods and results analysis are considered.

Main part

Linear regression theory

Linear regression is a mathematical model that predicts a dependent variable (y) based on independent variables (x) - this.

Linear regression works on the following mathematical formula:

$$= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + y \dots + \beta_n X_N + \varepsilon$$

Where: y is the predictive variable, x_i is the incoming factors, β_i is the coefficients, ε is the error.

Areas of application:

- Determination of house prices
- Economic forecasts
- Evaluation of results in scientific experiments
- Predicting supply and demand

Example of application:

Home price prediction

¹ Professor, of the Department of Computer Engineering and artificial intelligence, Fergana State Technical University

² Master of the Department of Computer Engineering and artificial intelligence, Fergana State Technical University



Factors: room number, Area, location rating.

$R^2 = 0.89$).

Predicting the cost of housing

Housing data in California, Katta. The goal is to predict home prices (price) using independent variables.

Software: Python environment and the following libraries:

```
import pandas as PD
```

```
import numpy as NP
```

```
import matplotlib as pyplot plt
```

```
import sea horn as SNS
```

```
from sklearn.model_selection import train_test_split
```

```
from sklearn.linear_model import LinearRegression
```

```
from sklearn.metrics import mean_squared_error, r2_score
```

Model evaluation criteria: $MSE \approx 0.53$, $R^2 \approx 0.60$

Model evaluation criteria

MSE (mean quadratic error) - dos.

R^2 score-model accuracy

Accuracy - determination of accuracy (determination of accuracy)

Confusion matrix

Accuracy and recall-Ozrtxbb. face

Results and analysis

Earnings between, width, length.

Conclusion

According to the results of the study, the linear regression model shows high efficiency in predicting nonlinear values such as price, income, score. Linear regression is effective in predicting numerical values. These approaches can be applied in economic, marketing and analytical matters.

Literature used

1. James, G., Witten, D., Hastie, T., Tibshirani, R. introduction to statistical education. Springer, 2021.
2. Bishop, C. M. pattern recognition and machine learning. Springer, 2006.
3. Scikit-learn: <https://scikit-learn.org>
4. Khalilov, D., Bozorova, S., Khonturaev, S., ... Sotvoldieva, D., Toshmatov, S. Self-learning system and methods of selection of weight coefficients of neural network. E3S Web of Conferences, 2024, 508, 04011

