Optimization and Machine Learning M

Laurea Magistrale in Telecommunications engineering
Prof. Andrea Lodi

Machine Learning Project 15 Jun 2023

Objective:
The objective of this project is to develop a machine learning model that can predict the salary of an individual based on relevant features such as years of experience, education level, and other factors using the "Salary_Data" dataset.
Tasks:

1. Data Exploration and Preprocessing:
   (a) Load the "Salary_Data" dataset.
   (b) Handle any missing values or outliers, if present.
   (c) Visualize the distribution of the "YearsExperience" and "Salary" variables using appropriate plots (e.g., histograms, scatter plots).
2. Feature Selection and Engineering:
   (a) Identify the relevant features from the dataset that can potentially influence salary prediction.
   (b) Perform feature engineering, if necessary, such as scaling or transforming features.
3. Model Development and Evaluation:
   (a) Split the dataset into training and testing subsets.
   (b) Choose an appropriate machine learning algorithm (e.g., linear regression, decision tree regression, or random forest regression) to train the model.
   (c) Train the model using the training data and evaluate its performance using suitable evaluation metrics (e.g., mean squared error, mean absolute error, R-squared score).
   (d) Fine-tune the model, if required, by adjusting hyperparameters to improve performance.
4. Prediction and Interpretation:
   (a) Use the trained model to make predictions on the testing data.
   (b) Assess the model's performance by comparing the predicted salaries with the actual salaries.
   (c) Interpret the results and discuss the findings.
5. Conclusion and Recommendations:
   (a) Summarize the project, highlighting the key steps and outcomes.
   (b) Provide recommendations for future improvements or possible extensions to the project.

Note: Throughout the project, document your code, explain the rationale behind your choices, and present the results in a clear and organized manner. Additionally, include visualizations, tables, and relevant metrics to support your analysis.