Coding Blocks | Online

Machine Learning Online

Assignment - 3 Linear and Logistic Regression

Part-I: Linear Regression

Objective:

To build a foundational understanding of linear regression for predictive modeling, learn model training and evaluation, and interpret regression outputs.

Assignment Tasks

Task 1: Simple Linear Regression

- Use the `scikit-learn` library.
- Load the **California Housing Dataset** (or any house price dataset).
- Predict the house price using a single feature (e.g., number of rooms).
- Train a simple linear regression model.
- Plot the regression line using `matplotlib`.

Task 2: Multiple Linear Regression

- Use the same dataset or load another dataset with at least 4 numeric features.
- Predict the target variable using **multiple independent variables**.
- Evaluate the model using:
    - R-squared
    - Mean Squared Error (MSE)
    - Root Mean Squared Error (RMSE)

- Display the coefficients of each feature.

Task 3: Feature Scaling and Normalization

- Demonstrate the effect of standardizing features using `StandardScaler`.
- Compare model performance **before and after scaling**.

Task 4: Model Interpretation

- Write Python code to:
    - Show the correlation matrix using a heatmap.
    - Identify which features have the **strongest relationship** with the target.
    - Discuss multicollinearity and how it might affect regression.

Part-II: Logistic Regression

Objective:

To understand classification problems, build logistic regression models, evaluate model accuracy, and interpret probabilities and decision boundaries.

Assignment Tasks

Task 5: Binary Classification with Logistic Regression

- Load the **Breast Cancer Dataset** from `sklearn.datasets`.
- Train a logistic regression model to predict whether a tumor is **benign or malignant**.
- Evaluate using:
    - Accuracy
    - Confusion Matrix
    - Precision, Recall, F1-score
    - ROC-AUC curve

Task 6: Threshold Tuning and Probability Interpretation

- Predict the probability of the positive class.
- Change the threshold (e.g., 0.3, 0.5, 0.7) and observe how the confusion matrix and F1-score change.
- Plot the ROC curve and mark optimal threshold.

Task 7: Multiclass Classification (Optional)

- Use the **Iris Dataset** to perform multiclass logistic regression.
- Use one-vs-rest or multinomial strategy.
- Evaluate model accuracy and classification report.

Part-III: General Questions

Objective:

To strengthen conceptual understanding of the differences, use cases, and assumptions of linear vs. logistic regression.

Assignment Tasks

Answer the following in brief (2–4 sentences each):

1. What are the assumptions of linear regression?
2. When should you use logistic regression instead of linear regression?
3. What is the interpretation of coefficients in logistic regression?
4. What is the difference between `sigmoid` and `softmax` functions?
5. Why is R-squared not suitable for evaluating logistic regression models?

Coding Blocks | Online