Milestone 2 – Documentation

1. Introduction

In this milestone, I learned and implemented multiple regression concepts to predict the Fire Weather Index (FWI) using the Algerian Forest Fire dataset. The key concepts learned include linear regression, ridge regression, sparsity, evaluation metrics (MSE and MAE), regularization parameter alpha, and saving models using .pkl files.

2. Linear Regression

Linear Regression models the relationship between independent variables (features) and a continuous target variable by fitting a straight-line equation. It is easy to interpret and works well when features are not highly correlated. However, it is sensitive to multicollinearity and may overfit in complex datasets.

3. Ridge Regression

Ridge Regression is an improved version of Linear Regression that adds L2 regularization to reduce overfitting and stabilize coefficients. It is especially useful when the dataset contains multicollinearity. The Ridge loss function includes a penalty term: MSE + alpha * $\Sigma(w^2)$, which shrinks coefficients and prevents them from becoming too large.

4. Sparsity

Sparsity refers to having many zero-valued coefficients in a model. Sparse models are easy to interpret, require less memory, and reduce overfitting. Lasso regression produces sparse models, whereas Ridge regression does not set coefficients to zero but only shrinks them.

5. Alpha (Regularization Strength)

Alpha is a hyperparameter that controls the amount of regularization applied in Ridge Regression. A small alpha behaves like Linear Regression, a moderate alpha improves generalization, and a very large alpha may lead to underfitting. RidgeCV automatically selects the best alpha using cross-validation.

6. Evaluation Metrics

Mean Squared Error (MSE): Measures the average squared difference between predicted and actual values. Lower MSE indicates better performance.
Mean Absolute Error (MAE): Measures the average absolute error. It is easier to interpret and less sensitive to outliers compared to MSE.

7. Saving the Model as a .pkl File

Machine learning models can be saved using joblib, allowing reuse without retraining. Saving: joblib.dump(model, 'model.pkl') Loading: model = joblib.load('model.pkl')

8. Key Learnings of the Week

- Linear and Ridge Regression understanding
- Sparsity concept and importance
- Working with MSE and MAE metrics
- Hyperparameter tuning using alpha
- RidgeCV for automatic alpha selection
- Saving ML models using .pkl files