Computer Science & Information Systems

Machine Learning Lab sheet - Module 1

EXERCISE 1 – END TO END ML PROJECT

1 Objective

The objective is to

- Go through an example ML project end to end

2 Steps to be performed

Tool	-	Python3, Google Colaboratory
Libraries required	-	numpy, pandas, matplotlib, sklearn
Input	-	the California Housing Prices dataset from the StatLib repository
ML Models	-	Linear Regression, Decision Trees, Random Forests, Support Vector Machines
Implementation	-	01_end_to_end_machine_learning_project.ipynb

Steps -

- Get the data.
    - Download the data
    - Explore the data structure
    - Create a Test Set
- Discover and visualize the data to gain insights.
    - Visualizing geographical data
    - Looking for correlations
    - Experimenting with Attribute combinations
- Prepare the data for Machine Learning algorithms.
    - Data Cleaning
    - Handling text and categorical attributes
    - Custom Transformers

    - Feature Scaling
    - Transformation Pipelines
- Select and Train a Model
    - Training and Evaluating on the Training Set
    - Better Evaluation Using Cross-Validation
- Fine-tune your Model
    - Grid Search
    - Randomized Search
    - Analyze the Best Models and Their Errors
    - Evaluate Your System on the Test Set

3 Expected Results

- Evaluate the Root Mean Square Errors for the models

4 Observation

- The data was thoroughly understood and prepared for ML algorithms
- ML models were trained, evaluated and fine-tuned

5 Modifications

- Try a different dataset.
- Try adding a transformer in the preparation pipeline to select only the most important attributes.
- Try creating a single pipeline that does the full data preparation plus the final prediction.
- Automatically explore some preparation options using GridSearchCV.