Final Project Proposal

Yiming Hao

COGS 109

- Introduction:
    - Define the problem:
        - The goal is to predict the sale prices of residential homes in Ames, Iowa based on various features.
    - Importance and relevance:
        - The accurate prediction of house prices is crucial for buyers, sellers, and real estate professionals. It can aid in decision-making, investment analysis, and understanding housing market trends.
- Dataset information:
    - Dataset link: *https://www.kaggle.com/c/house-prices-advanced-regression-techniques/data*
    - The dataset comprises observations of residential homes in Ames, Iowa. It includes features such as number of bedrooms, lot size, bathrooms, and others. The target variable is the sale price of the houses. The dataset consists of n observations and relevant predictors that can influence house prices. Factors affecting data quality may include missing values, outliers, or skewed distributions.
- Hypothesis:
    - Hypothesis statement:
        - We hypothesize that the size of the lot and the number of bedrooms will have a significant impact on the sale prices of houses."
- Methods:

    - Data analysis approach:
        - The project focuses on a prediction task, aiming to build a regression model to predict house prices based on the provided features.
- Rationale for the chosen approach:
    - Explain why regression is appropriate for predicting house prices. Discuss how regression models can capture relationships between predictors and the target variable, enabling accurate price estimation.
- Models
    - We are going to try the following models on this predictive task:
        - Linear Regression
        - Decision Trees
        - Ridge Regression and Lasso Regression
    - We are going to compare the performance of the models using various techniques, such as MSE, accuracy, cross-validation, etc.