Laboratory Activities for Week 7: Regression

SC310005 Artificial Intelligence
Khon Kaen Business School

(10 Points) Housing Prices Dataset

Dataset Overview and Objectives

Dataset Description:

Dataset:
https://raw.githubusercontent.com/kaopanboonyuen/SC310005_ArtificialIntelligence_2023s1/main/dataset/house_prices_dataset.csv

The dataset provided contains detailed information about various aspects related to housing properties. It includes various features such as building characteristics, lot size, property condition, amenities, and sale details. The target variable, 'SalePrice', represents the property's sale price in dollars, and the dataset includes multiple other columns that can influence this target.

Motivation:
The motivation behind using this dataset is to explore and apply regression modelling techniques to predict house prices accurately. Understanding the key factors influencing house prices can provide valuable insights for real estate stakeholders, buyers, sellers, and investors.

Assignment Objectives

Objective:
The primary objective of this assignment is to create regression models using the House Prices dataset to predict property sale prices accurately.

Assignment Problem:

Task for Students:

- ☐ Form groups and utilise the provided House Prices dataset.
- ☐ Select **a minimum of 5 variables** (e.g., Feature1 to Feature5) to build regression models.
- ☐ Train regression models using appropriate techniques.
- ☐ Evaluate model performance using **Mean Absolute Error (MAE).**
- ☐ Summarize obtained results and **MAE values.**
- ☐ Derive and **explain the regression equation** for each group's model.

Submission Details:

- ☐ Each group must submit their regression model code, MAE values, and explanations of the derived regression equations.
- ☐ The group with the lowest MAE value will receive recognition as the Winner and gain additional points.

Data Dictionary:
The dataset comprises various columns that describe different aspects related to properties. Here is a brief explanation of some key variables:

- SalePrice: The property's sale price in dollars. This is the target variable that you're trying to predict.
- MSSubClass: The building class.
- MSZoning: The general zoning classification.

- LotFrontage: Linear feet of street connected to the property.
- LotArea: Lot size in square feet.
- Street: Type of road access.
- Alley: Type of alley access.
- LotShape: General shape of the property.
- LandContour: Flatness of the property.
- Utilities: Type of utilities available.
- LotConfig: Lot configuration.
- LandSlope: Slope of property.
- Neighborhood: Physical locations within Ames city limits.
- Condition1: Proximity to main road or railroad.
- Condition2: Proximity to main road or railroad (if a second is present).
- BldgType: Type of dwelling.
- HouseStyle: Style of dwelling.
- OverallQual: Overall material and finish quality.
- OverallCond: Overall condition rating.
- YearBuilt: Original construction date.
- YearRemodAdd: Remodel date.
- RoofStyle: Type of roof.
- RoofMatl: Roof material.
- Exterior1st: Exterior covering on house.
- Exterior2nd: Exterior covering on house (if more than one material).
- MasVnrType: Masonry veneer type.
- MasVnrArea: Masonry veneer area in square feet.
- ExterQual: Exterior material quality.
- ExterCond: Present condition of the material on the exterior.
- Foundation: Type of foundation.
- BsmtQual: Height of the basement.
- BsmtCond: General condition of the basement.
- BsmtExposure: Walkout or garden level basement walls.
- BsmtFinType1: Quality of basement finished area.
- BsmtFinSF1: Type 1 finished square feet.
- BsmtFinType2: Quality of second finished area (if present).
- BsmtFinSF2: Type 2 finished square feet.
- BsmtUnfSF: Unfinished square feet of basement area.
- TotalBsmtSF: Total square feet of basement area.
- Heating: Type of heating.
- HeatingQC: Heating quality and condition.
- CentralAir: Central air conditioning.
- Electrical: Electrical system.
- 1stFlrSF: First-floor square feet.
- 2ndFlrSF: Second-floor square feet.
- LowQualFinSF: Low-quality finished square feet (all floors).
- GrLivArea: Above grade (ground) living area square feet.

- BsmtFullBath: Basement full bathrooms.
- BsmtHalfBath: Basement half bathrooms.
- FullBath: Full bathrooms above grade.
- HalfBath: Half baths above grade.
- Bedroom: Number of bedrooms above basement level.
- Kitchen: Number of kitchens.
- KitchenQual: Kitchen quality.
- TotRmsAbvGrd: Total rooms above grade (does not include bathrooms).
- Functional: Home functionality rating.
- Fireplaces: Number of fireplaces.
- FireplaceQu: Fireplace quality.
- GarageType: Garage location.
- GarageYrBlt: Year garage was built.
- GarageFinish: Interior finish of the garage.
- GarageCars: Size of garage in car capacity.
- GarageArea: Size of garage in square feet.
- GarageQual: Garage quality.
- GarageCond: Garage condition.
- PavedDrive: Paved driveway.
- WoodDeckSF: Wood deck area in square feet.
- OpenPorchSF: Open porch area in square feet.
- EnclosedPorch: Enclosed porch area in square feet.
- 3SsnPorch: Three-season porch area in square feet.
- ScreenPorch: Screen porch area in square feet.
- PoolArea: Pool area in square feet.
- PoolQC: Pool quality.
- Fence: Fence quality.
- MiscFeature: Miscellaneous feature not covered in other categories.
- MiscVal: Value of miscellaneous feature.
- MoSold: Month Sold.
- YrSold: Year Sold.
- SaleType: Type of sale.
- SaleCondition: Condition of sale.