**Lab 4.1.2**

**Interpretation of the Regression Evaluation Metrics**

1. **Mean Squared Error (MSE):**

* Measures the average of the squares of the errors, i.e., the average squared difference between the estimated values and the actual value.
* A lower MSE indicates a better fit.

**Expected Value Range:**

* **0 to ∞:** MSE values range from 0 to positive infinity.
* **Lower Values Indicate Better Fit:** The closer the MSE is to 0, the better the model's predictions match the actual values.

**Interpretation:**

* **Near 0:** Indicates very accurate predictions with minimal error.
* **High Values:** Suggests large errors between predicted and actual values, indicating a poorly fitting model.

**When to Use:** When large errors are particularly undesirable.  
**Example Scenario: Hospital Readmission Rates**

* **Context:** You are developing a model to predict the likelihood of patients being readmitted to a hospital within 30 days of discharge.
* **Use Case:** In healthcare, large prediction errors can have severe consequences, such as failing to provide necessary follow-up care for high-risk patients.
* **Why MSE:** MSE penalizes larger errors more heavily, ensuring that your model is sensitive to significant mispredictions. This can help prioritize interventions for those at highest risk of readmission.

1. **Root Mean Squared Error (RMSE):**

* Provides a measure of the average magnitude of the error.
* The RMSE gives an indication of how large the residuals are spread out. A lower RMSE indicates a better fit.

**Expected Value Range:**

* **0 to ∞:** RMSE values range from 0 to positive infinity.
* **Lower Values Indicate Better Fit:** Like MSE, lower RMSE values indicate a better fit. RMSE is in the same units as the target variable.

**Interpretation:**

* **Near 0:** Indicates very accurate predictions.
* **High Values:** Indicates larger prediction errors. The actual range depends on the specific context and units of the target variable.

**When to Use:** When you need an error metric that is in the same units as the predicted variable, making it easier to interpret.  
**Example Scenario: Weather Forecasting**

* **Context:** You are building a model to predict daily temperatures for a city.
* **Use Case:** Accurate temperature predictions are crucial for agriculture, event planning, and daily activities. The errors should be easily interpretable in the context of the actual temperature.
* **Why RMSE:** RMSE provides a measure of error in the same units as the predicted variable (degrees). This helps stakeholders understand how far off the predictions are from actual temperatures, making the metric more intuitive.

1. **Mean Absolute Error (MAE):**

* Measures the average magnitude of the errors in a set of predictions, without considering their direction.
* The MAE is the average over the test sample of the absolute differences between prediction and actual observation where all individual differences have equal weight. A lower MAE indicates a better fit.

**Expected Value Range:**

* **0 to ∞:** MAE values range from 0 to positive infinity.
* **Lower Values Indicate Better Fit:** Lower MAE values suggest more accurate predictions.

**Interpretation:**

* **Near 0:** Indicates very accurate predictions with minimal error.
* **High Values:** Indicates larger errors on average between predicted and actual values. The range will depend on the context and units of the target variable.

**When to Use:** When you need a straightforward average error that is less sensitive to outliers.  
**Example Scenario: Budget Forecasting for a Company**

* **Context:** You are predicting the quarterly budget requirements for different departments in a company.
* **Use Case:** You need a metric that provides a clear and interpretable average error in monetary terms without overemphasizing large deviations that might occur due to unexpected events.
* **Why MAE:** MAE gives an average error in the same units as the budget (dollars), making it straightforward to understand and communicate. It also ensures that each error contributes equally to the overall metric, which can be beneficial for financial planning.

1. **R-squared (R²):**

* Indicates how well the independent variables explain the variability of the dependent variable.
* An R² value of 0.75 means that 75% of the variance in the dependent variable is predictable from the independent variables.
* Higher R² values indicate a better fit, but it's important to consider the context and complexity of the model to avoid overfitting.

**Expected Value Range:**

* **0 to 1:** R² values range from 0 to 1. In some cases, R² can be negative if the model is performing worse than a horizontal line (mean of the dependent variable).

**Interpretation:**

* **Near 1 (e.g., 0.9 or higher):** Indicates that the model explains most of the variance in the dependent variable. This is a sign of a good fit.
* **Between 0.5 and 0.9:** Indicates a moderate fit where the model explains a significant portion of the variance.
* **Near 0 (e.g., 0.1 or lower):** Indicates that the model explains very little of the variance, suggesting a poor fit.
* **Negative Values:** Indicate that the model is performing worse than the mean model, which can happen if the chosen model does not fit the data well.

**When to Use:** When you want to assess how well your model explains the variability of the dependent variable.  
**Example Scenario: Evaluating Marketing Campaigns**

* **Context:** You are analysing the effectiveness of various marketing campaigns in driving sales.
* **Use Case:** You want to know how well the independent variables (e.g., ad spend, campaign duration, target demographics) explain the variability in sales figures.

**Why R²:** R-squared provides a measure of how well your independent variables explain the variability in the dependent variable (sales). A higher R² indicates that your model captures a significant portion of the variance, helping you understand the impact of different marketing strategies.

Plotting a diagonal correlation matrix:

https://seaborn.pydata.org/examples/many\_pairwise\_correlations.html