The Linear Regression Model: Takeaways

Syntax

- Importing and instantiating a linear regression model:

```
from sklearn.linear_model import LinearRegression
lr = LinearRegression()
```

- Using the

```
LinearRegression
```

class to fit a linear regression model between a set of columns:

```
lr.fit(train[['Gr Liv Area']], train['SalePrice'])
```

- Returning the and parameters for :

```
a0 = lr.intercept_
a1 = lr.coef_
```

- Predicting the labels using the training data:

```
test_predictions = lr.predict(test[['Gr Liv Area']])
```

- Calculating the correlation between pairs of columns:

```
train[['Garage Area', 'Gr Liv Area', 'Overall Cond', 'SalePrice']].corr()
```

Concepts

- An instance-based learning algorithm, such as K-nearest neighbors, relies completely on previous instances to make predictions. K-nearest neighbors doesn't try to understand or capture the relationship between the feature columns and the target column.

- Parametric machine learning, like linear regression and logistic regression, results in a mathematical function that best approximates the patterns in the training set. In machine learning, this function is often referred to as a model. Parametric machine learning approaches work by making assumptions about the relationship between the features and the target column.
- The following equation is the general form of the simple linear regression model:

  where represents the target column while represents the feature column we chose to use in our model. and represent the parameter values that are specific to the dataset.
- The goal of simple linear regression is to find the optimal parameter values that best describe the relationship between the feature column and the target column.
- We minimize the model's residual sum of squares to find the optimal parameters for a linear regression model. The equation for the model's residual sum of squares is as follows:

  where is our target column and are our true values.
- A multiple linear regression model allows us to capture the relationship between multiple feature columns and the target column. The formula for multiple linear regression is as follows:

  where to are our feature columns, and the parameter values that are specific to the data set are represented by along with to .
- In linear regression, it is a good idea to select features that are a good predictor of the target column.

Resources

- Linear Regression Documentation
- pandas.DataFrane.corr() Documentation