Summary/Review

Introduction to Supervised Machine Learning

The types of supervised Machine Learning are:

- Regression, in which the target variable is continuous
- Classification, in which the target variable is categorical

To build a classification model you need:

- Features that can be quantified
- A labeled target or outcome variable
- Method to measure similarity

Linear Regression

A linear regression models the relationship between a continuous variable and one or more scaled variables.It is usually represented as a dependent function equal to the sum of a coefficient plus scaling factors times the independent variables.

Residuals are defined as the difference between an actual value and a predicted value.

A modeling best practice for linear regression is:

- Use cost function to fit the linear regression model
- Develop multiple models
- Compare the results and choose the one that fits your data and whether you are using your model for prediction or interpretation.

Three common measures of error for linear regressions are:

- Sum of squared Error (SSE)
- Total Sum of Squares (TSS)
- Coefficient of Determination (R2)

Linear Regression Syntax

The most simple syntax to train a linear regression using scikit learn is:

```
from sklearn.linear_model import LinearRegression

LR = LinearRegression()

LR = LR.fit(X_train, y_train)
```

To score a data frame X_test you would use this syntax:

```
y_predict = LR.predict(X_test)
```