ACTIVIDAD #3

Tipo actividad: Reading comprehension: "Regression and Classification | Supervised Machine Learning" , matching exercise about the previous reading

9) Socialization key vocabulary reading #3: "Regression and Classification | Supervised Machine Learning"

- **Function Approximation:** Meaning: The process of developing a model using historical data to predict outcomes in new data where the answer is unknown. It involves approximating a mapping function from input variables to output variables.
- **Classification Predictive Modeling:** Meaning: The task of approximating a mapping function from input variables to discrete output variables. It involves predicting the class or category for a given observation, with output variables referred to as labels or categories.
- **Regression Predictive Modeling:** Meaning: The task of approximating a mapping function from input variables to a continuous output variable. It involves predicting a quantity, typically a real-value, such as an integer or floating point value.

- **Classification Accuracy:** Meaning: A measure of the percentage of correctly classified examples out of all predictions made by a classification model. It evaluates the accuracy of predictions for discrete output variables.
- **Root Mean Squared Error (RMSE):** Meaning: A common metric used to estimate the skill of a regression predictive model. It measures the average magnitude of the errors between predicted and actual values, providing a sense of the model's accuracy.
- **Multivariate Regression Problem:** Meaning: A regression problem with multiple input variables. It involves predicting a continuous output variable based on multiple input variables.
- **Classification vs Regression:** Meaning: A comparison between classification and regression predictive modeling problems. Classification predicts discrete class labels, while regression predicts continuous quantities. The evaluation metrics for these tasks, such as accuracy and RMSE, are distinct and do not overlap.

10) Reading comprehension #3: "Regression and Classification | Supervised Machine Learning"

Reading: "Regression and Classification | Supervised Machine Learning"

Function Approximation

Predictive modeling is the problem of developing a model using historical data to make a prediction on new data where we do not have the answer.

Predictive modeling can be described as the mathematical problem of approximating a mapping function (f) from input variables (X) to output variables (y). This is called the problem of function approximation.

The job of the modeling algorithm is to find the best mapping function we can given the time and resources available.

Generally, we can divide all function approximation tasks into classification tasks and regression tasks.

uClassification Predictive Modeling

Classification predictive modeling is the task of approximating a mapping function (f) from input variables (X) to discrete output variables (y).

The output variables are often called labels or categories. The mapping function predicts the class or category for a given observation.

For example, an email of text can be classified as belonging to one of two classes: “spam“ and “not spam“.

A classification problem requires that examples be classified into one of two or more classes.

A classification can have real-valued or discrete input variables.

A problem with two classes is often called a two-class or binary classification problem.

A problem with more than two classes is often called a multi-class classification problem.

A problem where an example is assigned multiple classes is called a multi-label classification problem.

It is common for classification models to predict a continuous value as the probability of a given example belonging to each output class. The probabilities can be interpreted as the likelihood or confidence of a given example belonging to each class. A predicted probability can be converted into a class value by selecting the class label that has the highest probability.

For example, a specific email of text may be assigned the probabilities of 0.1 as being “spam” and 0.9 as being “not spam”. We can convert these probabilities to a class label by selecting the “not spam” label as it has the highest predicted likelihood.

There are many ways to estimate the skill of a classification predictive model, but perhaps the most common is to calculate the classification accuracy.

The classification accuracy is the percentage of correctly classified examples out of all predictions made.

Regression Predictive Modeling

Regression predictive modeling is the task of approximating a mapping function (f) from input variables (X) to a continuous output variable (y).

A continuous output variable is a real-value, such as an integer or floating point value. These are often quantities, such as amounts and sizes.

For example, a house may be predicted to sell for a specific dollar value, perhaps in the range of 200,000.

A regression problem requires the prediction of a quantity.

A regression can have real valued or discrete input variables.

A problem with multiple input variables is often called a multivariate regression problem.

A regression problem where input variables are ordered by time is called a time series forecasting problem.

Because a regression predictive model predicts a quantity, the skill of the model must be reported as an error in those predictions.

There are many ways to estimate the skill of a regression predictive model, but perhaps the most common is to calculate the root mean squared error, abbreviated by the acronym RMSE.

A benefit of RMSE is that the units of the error score are in the same units as the predicted value.

An algorithm that is capable of learning a regression predictive model is called a regression algorithm.

Some algorithms have the word "regression" in their name, such as linear regression and logistic regression, which can make things confusing because linear regression is a regression algorithm whereas logistic regression is a classification algorithm.

Classification vs Regression

Classification predictive modeling problems are different from regression predictive modeling problems.

Classification is the task of predicting a discrete class label.

Regression is the task of predicting a continuous quantity.

There is some overlap between the algorithms for classification and regression; for example:

A classification algorithm may predict a continuous value, but the continuous value is in the form of a probability for a class label.

A regression algorithm may predict a discrete value, but the discrete value in the form of an integer quantity.

Some algorithms can be used for both classification and regression with small modifications, such as decision trees and artificial neural networks. Some algorithms cannot, or cannot easily be used for both problem types, such as linear regression for regression predictive modeling and logistic regression for classification predictive modeling.

Importantly, the way that we evaluate classification and regression predictions varies and does not overlap, for example:

Classification predictions can be evaluated using accuracy, whereas regression predictions cannot.

Regression predictions can be evaluated using root mean squared error, whereas classification predictions cannot.

Adapted from: https://machinelearningmastery.com/classification-versus-regression-in-machine-learning/

11) Matching heading definition.

Match the definitions with the appropriate heading:

Column A: Headings

1. Function Approximation
2. Classification Predictive Modeling
3. Regression Predictive Modeling
4. Classification Accuracy
5. Root Mean Squared Error (RMSE)

Column B: Definitions

A. The task of approximating a mapping function from input variables to discrete output variables, predicting the class or category for a given observation, often referred to as labels or categories.

B. The problem of developing a model using historical data to predict outcomes in new data where the answer is unknown, involving the approximation of a mapping function from input variables to output variables.

C. A common metric used to estimate the skill of a regression predictive model, measuring the average magnitude of errors between predicted and actual values, providing a sense of the model's accuracy.

D. The task of approximating a mapping function from input variables to a continuous output variable, predicting a quantity, typically a real-value, such as an integer or floating point value.

E. A measure of the percentage of correctly classified examples out of all predictions made by a classification model, evaluating the accuracy of predictions for discrete output variables.