Topic 4: Support Vector Machine Regressor

Jonathan Ibifubara Pollyn

GitHub Repository: https://github.com/JonathanPollyn/Machine-Learning-for-Data-Science

College of Science, Engineering and Technology, Grand Canyon University

DSC-540: Machine Learning for Data Science

Dr. Aiman Darwiche

November 24, 2021

Number of misclassifications as the performance measure and maximizing the margin between the two classes as the performance measure

Given that the training dataset is simply a rough approximation of the rest of the data universe, such solutions may result in misclassifications of points that aren't visible. Some of the elements that will influence my decisions are as follows: When the dataset is small, the hyperplane clearly distinguishes both groups with a large margin. In addition, it works well in n-dimensional spaces, and the number of features exceeds the number of data points in this case. Finally, the importance of memory cannot be overstated.

SVM can be used for classification. They can be applied to both linear and non-linear problems. SVM algorithm tries to separate one class data point from the other class data points using a separating hyperplane such that both the classes are very far from the plane. The separating plane of both the classes is called the hyperplane. SVM can handle any complex relationship between both classes using the kernel trick. The drawback of this algorithm is that it is computationally expensive.

Binary classification task using (i) *linear SVM*, and (ii) *perceptron algorithm*

When your data contains precisely two classifications, you can use a support vector machine (SVM). The optimum hyperplane that separates all data points of one class from those of the other class is found by an SVM. For an SVM, the best hyperplane is the one that has the most significant margin between the two classes. The maximum width of the slab parallel to the hyperplane with no inside data points is the margin. The support vectors are the data points nearest to the separating hyperplane, located on the slab's edge. For linear SVM, SVM works robustly because SVM focuses on maximizing the margin. All the points are correctly classified and such that the number of misclassifications is minimized. The SVM margin helps us better accommodate the test data from each class and helps avoid overfitting or underfitting the data points. SVM works better with linear data because it looks to maximize the margin.

A linear classification algorithm is the Perceptron, which implies it learns a decision boundary in the feature space that separates two classes using a line (called a hyperplane). As a result, it's best for issues where the classes can be easily separated using a line or linear model, sometimes known as linearly separable problems. The stochastic gradient descent optimization procedure is used to train the model's coefficients, referred to as input weights. When samples are non-separable, it develops cycles. Every time an observation is visited, it responds. As a result, when data separation is not evident, it takes longer. The perceptron algorithm aims to cut down on errors.

Application

Regressor helps in establishing a relationship among the variables by estimating how one variable affects the other. Machine learning consists of mathematical methods that allow data scientists to predict a continuous outcome (y) based on the value of one or more predictor variables (x). The dependent variable in the testing dataset can be predicted using this freshly built model. Then, using accuracy measurements such as R-squared, root mean square error, root mean average error, Pearson correlation coefficient, and others, predicted values can be compared to the original dependent variable values. The percentage of datasets allotted to the training and testing datasets can be modified if the accuracy score isn't accurate enough and desired a stronger model. For example, suppose the training dataset accounted for 70% of the total dataset and the testing dataset for 30%. In that case, the training dataset can now account for 80% of the total dataset and the testing dataset for 20%.

It helps find the correlation between variables and predicts the continuous output variable based on one or more predictor variables. It is mainly used for prediction, forecasting, time series modeling, and determining the causal-effect relationship between variables. Some statistical metrics that are used to quantify its accuracy are Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Coefficient of determination or R2 and Adjusted R2. One may be forced to compromise accuracy by limiting the search space, conducting a sparser search, or doing both. A d-dimensional data collection with m-many maxima in each dimension, for example, would have a total of md maxima. A thorough search across all maxima for large d would be intractable.

References List

Brownlee, J. (2020, December 10). *Perceptron algorithm for classification in python*. Machine Learning Mastery. https://machinelearningmastery.com/perceptron-algorithm-for-classification-in-python/

Ganiyu, M. (2019, December 1). *How regression analysis works. different forms of regression analysis… | by mubarak ganiyu | towards data science*. Medium. https://towardsdatascience.com/how-regression-analysis-works-10f44c37b20a

Gopal, M. (2019). *Applied machine learning* (1st ed.). McGraw-Hill Professional.

Nelson, J. D., Damper, R. I., Gunn, S. R., & Guo, B. (2009). A signal theory approach to support vector classification: The sinc kernel. *Neural Networks*, *22*(1), 49–57. https://doi.org/10.1016/j.neunet.2008.09.016

Pykes, K. (2021, August 6). *The difference between classification and regression in machine learning | by kurtis pykes | towards data science*. Medium. https://towardsdatascience.com/the-difference-between-classification-and-regression-in-machine-learning-4ccdb5b18fd3

(n.d.). https://www.mathworks.com/help/stats/support-vector-machines-for-binary-classification.html