Support Vector Machine: Regression

Beny Maulana Achsan · Follow
Published in IT Paragon · 4 min read · Dec 10, 2019

👏 7   💬

Holla guys! Welcome to the next article. In this part, we will discuss further details of supervised machine learning.
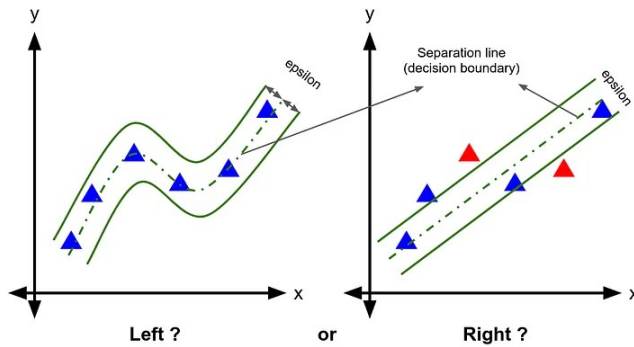

Figure 1: SVM Regression

Well, what do you think of the two pictures above? Which one is correct? The right or the left one? Don't worry, you will know the answer after reading this article :)
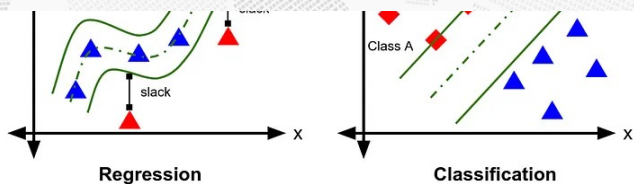
Figure 2: Regression vs Classification

Regression

If you want to predict continuous values, such as trying to predict the gold price, you would use regression. This type of example doesn't have a specific value constraint because the value could be any number with no limits.

Classification

If you want to predict discrete values, such as classifying something into two categories, you would use classification. A case like, "Will she make this

purchase" will have an answer that falls into two specific categories: yes or no. For the detail explanation, you can read this one "Support Vector Machine: Classification".

1. Regression

Regression is another form of supervised learning. As we have discussed above, the difference between classification and regression is that regression **outputs a number** rather than a class. Figure 1 shows two types of regression with the same epsilon. Well, both of them are correct. The right one tolerates some outlier points but the left one is trying to achieve zero tolerance with perfect regression. In a real-world application, finding the perfect regression for billions of training data sets takes a lot of time. As you will see in coding, regularization parameter and gamma should be defined. We can combine those parameters to achieve a considerable non-linear regression line with higher accuracy in a reasonable amount of time.

2. Tuning parameters : Regularization, Gamma, and Epsilon

Regularization

The **regularization** parameter (C parameter in python's sklearn library) tells the SVM optimization on how much you want to avoid misclassifying on each training sample. In other words:

> C *parameter balances the trade-off between the model complexity and empirical error. To simplify,* when C is large, the SVM tends to be overfitting while when C is small, the SVM tends to be underfitting.

Gamma

The **gamma** parameter defines how far the influence of a single training example reaches (low values mean far and a high value means close). With low gamma, points far away from plausible separation lines are considered in the calculation for the separation line. Whereas high gamma means the points close to a plausible line are considered in the calculation.

> When γ is large, the SVM tends to be overfitting. On the other hand, when γ is small, the SVM tends to be underfitting.

Epsilon

In the same way, as with a classification approach, there is a motivation to seek and optimize the generalization bounds given for regression. They relied on defining the loss function that ignores errors, which are situated within a certain distance of the true value. This type of function is often called *epsilon intensive loss function*.

> ε-*intensive loss function affects the smoothness of the SVM's response.*

The figure below shows an example of linear and non-linear regression function with *epsilon intensive band*. The variables measure the cost of the errors on the training points. These are zero for all blue points that are inside the band.
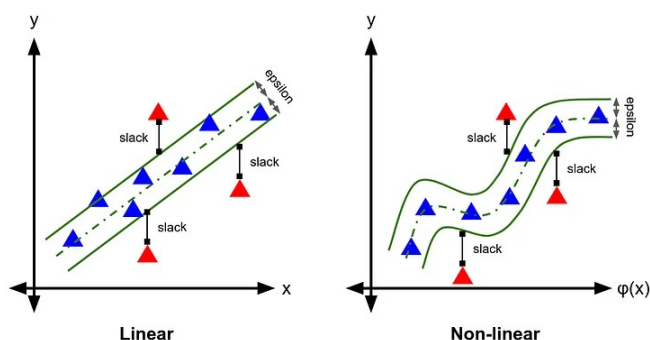

Figure 3: Linear (left) and Non-linear (right) Regression with Epsilon Intensive Band

In regression, epsilon has a function to give a tolerable error of the regression model. It depends on the problems. If we need a high tolerable error, we can increase the epsilon intensive band and vice versa. Now you know the answer to the question above. Thank you :)

TOGETHER WE DRIVE **TECHNOLOGY** AND **INNOVATION**

Machine Learning   Artificial Intelligence   Support Vector Machine   Regression

Big Data

7

Written by Beny Maulana Achsan

Follow

34 Followers · Editor for IT Paragon

@bmachsan

More from Beny Maulana Achsan and IT Paragon

Beny Maulana Achsan in IT Paragon

The 9 Most Powerful Digital Branding Strategies!

Recently, the term 'Digital Branding' has been popular on social media such as Instagram,...

16 min read · Mar 15, 2021

54   1

Nadia Saraswati in IT Paragon

The Cybersecurity Cube

Nowadays, everyone and everything uses secured and unsecured computer networks...

5 min read · Feb 21, 2021

17

Kurniawan Adi Saputra in IT Paragon

Bat != Bad: Brief Introduction to The Bat Algorithm

The work from home activity is due to the ongoing corona virus pandemic which is...

6 min read · Aug 23, 2020

11   2

Beny Maulana Achsan in IT Paragon

Strategic Management of Information System in The 21st...

Imagine now you are as a CIO in a corporate company. At a BOD meeting, your CEO Grou...

7 min read · Mar 19, 2022

1

See all from Beny Maulana Achsan   See all from IT Paragon

Recommended from Medium

NANDINI VERMA

An Introduction to Support Vector Regression (SVR) in Machine...

Support Vector Regression (SVR) is a machine learning technique used for...

3 min read · Nov 3, 2023

26 ⤹ 🔖

AnalytixLabs

Random Forest Regression — How it Helps in Predictive Analytics?

Random forest regression is an invaluable tool in data science. It enables us to make...

11 min read · Dec 26, 2023

101 ⤹ 🔖

Lists

Predictive Modeling w/ Python
20 stories · 1103 saves

AI Regulation
6 stories · 415 saves

Natural Language Processing
1377 stories · 870 saves

Practical Guides to Machine Learning
10 stories · 1317 saves

Megha Natarajan

Intuition behind C and Gamma in Support Vector Machines: A Visu...

⭐ · 3 min read · Nov 9, 2023

104 ⤹ 🔖

Nehra Vikas

ADABOOST ALGORITHM

When it comes to ensemble models, there are two major techniques — bagging and...

4 min read · 1 day ago

⤹ 🔖

Thommaskevin

TinyML — Random Forest (Classifier and Regressor)

From mathematical foundations to practical implementation.

12 min read · Dec 20, 2023

54 💬 2 🔖

Gen. David L.

How to Implement the Random Forest Algorithm by Python

Introduction

5 min read · Dec 22, 2023

2 ⤹ 🔖

See more recommendations