L1 and L2 Regularization Methods

Anuja Nagpal · Follow

Published in Towards Data Science · 2 min read · Oct 13, 2017

Machine Learning

In my last post, I covered the introduction to Regularization in supervised learning models. In this post, let's go over some of the regularization techniques widely used and the key difference between those.

In order to create less complex (parsimonious) model when you have a large number of features in your dataset, some of the Regularization techniques used to address over-fitting and feature selection are:

1. **L1 Regularization**

2. **L2 Regularization**

A regression model that uses L1 regularization technique is called *Lasso Regression* and model which uses L2 is called *Ridge Regression.*

The key difference between these two is the penalty term.

Ridge regression adds "*squared magnitude*" of coefficient as penalty term to the loss function. Here the *highlighted* part represents L2 regularization element.

$$\sum_{i=1}^{n}(y_i - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} \beta_j^2$$

Cost function

Here, if *lambda* is zero then you can imagine we get back OLS. However, if *lambda* is very large then it will add too much weight and it will lead to

Lasso Regression (Least Absolute Shrinkage and Selection Operator) adds "*absolute value of magnitude*" of coefficient as penalty term to the loss function.

$$\sum_{i=1}^{n}(Y_i - \sum_{j=1}^{p} X_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} |\beta_j|$$

Cost function

Again, if *lambda* is zero then we will get back OLS whereas very large value will make coefficients zero hence it will under-fit.

The **key difference** between these techniques is that Lasso shrinks the less important feature's coefficient to zero thus, removing some feature altogether. So, this works well for **feature selection** in case we have a huge number of features.

Traditional methods like cross-validation, stepwise regression to handle overfitting and perform feature selection work well with a small set of features but these techniques are a great alternative when we are dealing with a large set of features.

Machine Learning   Data Science   Analytics   Data Analytics

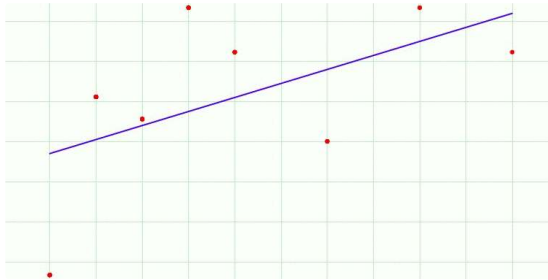intelligence professionals.

Written by Anuja Nagpal   Follow

2.5K Followers · 69 Following

Data Scientist/Machine Learning Engineer- Keeping up with Data Science and Machine Learning. #datascience #machinelearning #womenintech

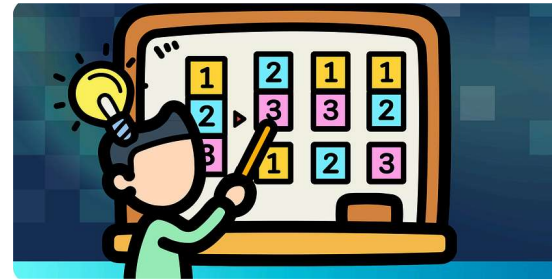## More from Anuja Nagpal and Towards Data Science

tds In Towards Data Science by Anuja Nagpal

Over-fitting and Regularization

Machine Learning
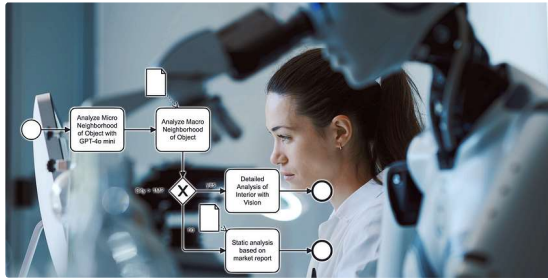
Oct 11, 2017   1.5K   10

tds In Towards Data Science by Samy Baladram

Model Validation Techniques, Explained: A Visual Guide with...

12 must-know methods to validate your machine learning

1d ago   125   3

tds In Towards Data Science by Dr. Marcel Müller

Why Internal Company Chatbots Fail and How to Use Generative AI...

Start with the problem and not with the solution

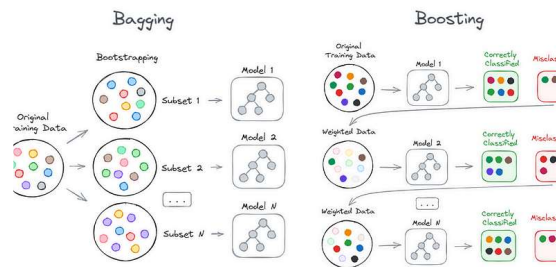2d ago   👏 450   💬 7

tds In Towards Data Science by Anuja Nagpal

Decision Tree Ensembles- Bagging and Boosting

Random Forest and Gradient Boosting

Oct 18, 2017   👏 1.96K   💬 10

See all from Anuja Nagpal   See all from Towards Data Science
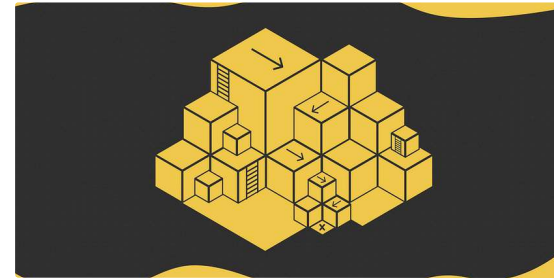
Recommended from Medium

In Towards AI by Thomas A Dorfer

Bagging vs. Boosting: The Power of Ensemble Methods in Machine...

How to maximize predictive performance by creating a strong learner from multiple weak...

Jun 16, 2023 · 👏 580 · 💬 4

In Towards Data Science by Vyacheslav Efimov

Understanding Deep Learning Optimizers: Momentum, AdaGra...

Gain intuition behind acceleration training techniques in neural networks

Dec 30, 2023 · 👏 456 · 💬 4

Lists

Predictive Modeling w/ Python

20 stories · 1699 saves

Practical Guides to Machine Learning

10 stories · 2064 saves

Natural Language Processing

1841 stories · 1463 saves

data science and AI

40 stories · 295 saves

In **Stackademic** by **Abdur Rahman**

Python is No More The King of Data Science

5 Reasons Why Python is Losing Its Crown

Oct 23   8.7K   34

Emmanuel Ikogho

Data Science is dying; here's why

Why 85% of data science projects fail

Sep 3   1.98K   97

Jessica Stillman

Jeff Bezos Says the 1-Hour Rule Makes Him Smarter. New...

Jeff Bezos's morning routine has long included the one-hour rule. New...

Oct 30   13.6K   315

Mark Shrime, MD, PhD

The dumbest decision I ever made (and the Nobel Prize that explains...

Decision science, Family Guy, and why the "safe" choice is often the riskiest.

Nov 21   6.8K   152

See more recommendations