

**Tom Ron**

Engineering Manager, Software Engineer, Data Scientist

# Davies-Bouldin Index

👤 tom   📁 Uncategorized   🕒 November 30, 2016   ⌵ 1 Minute  
TL;DR – Yet another clustering evaluation metric

Davies-Bouldin index was suggested by David L. Davies and Donald W. Bouldin in “A Cluster Separation Measure” (*IEEE Transactions on Pattern Analysis and Machine Intelligence*. PAMI-1 (2): 224–227. doi:10.1109/TPAMI.1979.4766909 (<https://dx.doi.org/10.1109%2FTPAMI.1979.4766909>), [full pdf \(https://www.researchgate.net/profile/Don\\_Bouldin/publication/224377470\\_A\\_Cluster\\_Separation\\_Measure/links/02e7e52d57df1ad121000000.pdf\)](https://www.researchgate.net/profile/Don_Bouldin/publication/224377470_A_Cluster_Separation_Measure/links/02e7e52d57df1ad121000000.pdf)))

Just like Silhouette score, Calinski-Harabasz index and Dunn index, Davies-Bouldin index provide an internal evaluation schema. I.e. the score is based on the cluster itself and not on external knowledge such as labels.

The Silhouette score reflects how similar a point is to the cluster it is associated with. I.e. for each point with compute the average distance of the point from the points in the nearest cluster minus the average distance of the point from the points in its own cluster divided by the maximum between those distances. The overall score is the average of the score per point. The Silhouette score is bounded from -1 to 1 and higher score means more distinct clusters.

The Calinski-Harabasz index compares the variance between-clusters to the variance within each cluster. This measure is much simpler to calculate than the Silhouette score however it is not bounded. The higher the score the better the separation is.

The intuition behind Davies-Bouldin index is the ratio between the within cluster distances and the between cluster distances and computing the average overall the clusters. It is therefore relatively simple to compute, bounded – 0 to 1, lower score is better. However, since it measures the distance between clusters' centroids it is restricted to using Euclidean distance function.

Silhouette score and Calinski-Harabasz index were previously implemented and are part of [scikit-learn](http://scikit-learn.org/stable/) (<http://scikit-learn.org/stable/>), and I implemented Davies-Bouldin index. Hopefully this will be my first contribution to scikit-learn, my implementation is [here](https://github.com/tomron/scikit-learn/tree/davies_bouldin_index) ([https://github.com/tomron/scikit-learn/tree/davies\\_bouldin\\_index](https://github.com/tomron/scikit-learn/tree/davies_bouldin_index)).

## Published by tom



[View all posts by tom](#)

## 6 thoughts on “Davies-Bouldin Index”

**jeff** says:

February 8, 2017 at 3:16 am

thanks for this nice implementation.

↪ Reply

**jain82mahak** says:

May 24, 2017 at 2:09 pm

Hey! It is a nice implementation. I copied your function `davies_bouldin_index` in my source code and imported required modules but I am getting a score greater than 1. Is it possible?

↪ Reply

**tom** says:

May 25, 2017 at 5:08 pm

I think that if the clustering is something like the result of the `dbscan` in the first row (here – [http://ogrisel.github.io/scikit-learn.org/sklearn-tutorial/auto\\_examples/cluster/plot\\_cluster\\_comparison.html](http://ogrisel.github.io/scikit-learn.org/sklearn-tutorial/auto_examples/cluster/plot_cluster_comparison.html)) than the centroids ( $M_{ij}$ ) are really close to each other while the  $S_i$ 's are larger and that can result values greater than 1.

↪ Reply

**Pedro** says:

October 12, 2017 at 3:16 am

Hi Tom! Congrats. I found your implementation by chance only after I implemented my own function to compute the DB index. I checked mine against yours and we're getting the same results (values greater than 1 included). Why is your implementation not merged to scikit-learn already on github? Could you check if there is something else you need to address? I would really like to see it in scikit. 😊

↪ Reply

**tom** says:

October 12, 2017 at 10:00 am

Thank you for the kind words.

This is the current status of the PR – <https://github.com/scikit-learn/scikit-learn/pull/7942>

I might get to it someday 😊

↪ Reply

**Pedro** says:

October 12, 2017 at 6:34 pm

I know, I downloaded the code from there. However, it's been stalled for a while now and it would be nice to finally merge it for us...

[Create a free website or blog at WordPress.com.](#) [Do Not Sell My Personal Information](#)

## Privacy & Cookies

We, WordPress.com, and our advertising partners store and/or access information on your device and also process personal data, like unique identifiers, browsing activity, and other standard information sent by your device including your IP address. This information is collected over time and used for personalised ads, ad measurement, audience insights, and product development specific to our ads program. If this sounds good to you, select "I Agree!" below. Otherwise, you can get more information, customize your consent preferences, or decline consent by selecting "Learn More". Note that your preferences apply to all websites in the [WordPress.com network](#), and if you change your mind in the future you can update your preferences anytime by visiting the Privacy link displayed under each ad. One last thing,