8/12/2021

















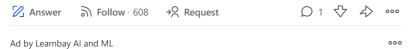
Add auestion

Quora uses cookies to improve your experience. Read More

L1 Regularization Regularization L2 Overfitting (modeling) Regression (statistics)

Data Science +2

# What is the difference between L1 and L2 regularization? How does it solve the problem of overfitting? Which regularizer to use and when?



### Would data science and AI be the right career option for me? Can I get my profile reviewed by an expert?

I have 9+ years of industry experience in MNCs in multiple domains and also helped 4000+ working professionals in career transitions into data science and Al in last 5 years. Start...(Continue reading in feed)

#### 29 Answers



**Manjari Narayan**, Ph.D. Electrical Engineering & Statistics, Rice University (2015)



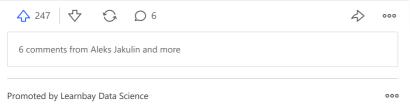
Updated November 16, 2014  $\cdot$  Upvoted by Anant Raj, Ph.D. Machine Learning & Mathematical Optimization, Max Planck Institute for Intelligent Systems (2020) and Justin Rising, MSE in CS, PhD in Statistics

Originally Answered: What is the difference between L1 and L2 regularization?

Justin Solomon has a great answer on the difference between L1 and L2 norms and the implications for regularization.

#### ℓ1 vs ℓ2 for signal estimation:

Here is what a signal that is sparse or approximately sparse i.e. that belongs to the ell-1 ball looks like. It becomes extremely unlikely that an  $\ell 2$  penalty can recover a sparse signal since very few solutions of such a cost function are truly sparse.  $\ell 1$  penalties on the other hand are great for recovering truly sparse signals, as they are computationally tractable but still capable of recovering the exact sparse solution.  $\ell 2$  penalization is pref... (more)



## Which is the best data science and AI certification for working professionals?



Krishna Kumar, Data science career coach and Founder at Learnbay (2015-present) Updated July 14, 2020

Best data science and Al Certification For Working Professionals: Learnbay offers IBM certified data science & Al courses which are designed for working professionals.

Based on your years of experience and your career goal, you (Continue reading in feed)



YunFang Juan, Applied Machine Learning to Yahoo! Search and Facebook

Updated September 29, 2013 · Upvoted by Alberto Bietti, PhD student in machine learning. Former ML engineer and Benjamin Golub, PhD in Economics from Stanford; I work in applied probability on statistical problems related to social netwo...

Originally Answered: What is the difference between L1 and L2 regularization?

Practically, I think the biggest reasons for regularization are 1) to avoid overfitting by not generating high coefficients for predictors that are sparse. 2) to stabilize the estimates especially when there's collinearity in the data.

1) is inherent in the regularization framework. Since there are two forces pulling each other in the objective function, if there's no meaningful loss reduction, the increased penalty from the regularization term wouldn't improve the overall objective function.

#### **Related Questions**

Why is L1-regularization better than L2-regularization, provided that all Norms are...

What is an intuitive explanation of regularization?

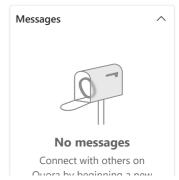
What is the advantage of combining L2 and L1 regularizations?

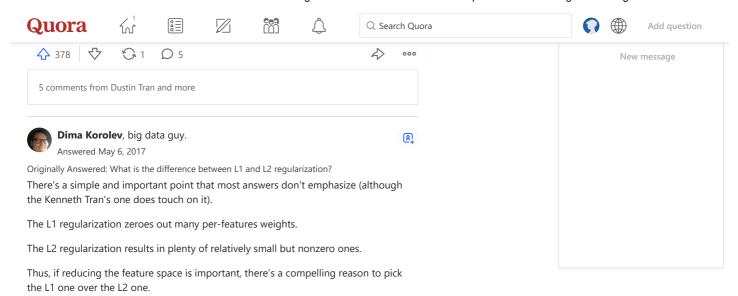
When is Ridge regression favorable over Lasso regression?

Why is L1 regularization supposed to lead to sparsity than L2?

Are there any criteria to distinguish overfitting?

Ask question





Natural examples of where it could be valuable to reduce the feature space could be:

- · when it's important to maximize the QPS of applying the model,
- when it's important to minimize the size of the data file with feature values which



#### Are you a Software Engineer living outside the US?

US Software Developers jobs. Work remotely. No visa needed. Full time jobs. High salary.



Related Questions

More Answers Below

Why is L1-regularization better than L2-regularization, provided that all Norms are equivalent?

What is an intuitive explanation of regularization?

What is the advantage of combining L2 and L1 regularizations?



**Leihua Ye, Ph.D. Researcher**, Ph.D. Data Science & Political Science, University of California, Santa Barbara (2020)

Ø

Answered January 15

• Part 1: What They Are

LASSO regression, L1 regularization, includes a hyper-parameter  $\alpha$  times the sum of the absolute value of the coefficients as penalty term in its cost function, shown below (marked in red):

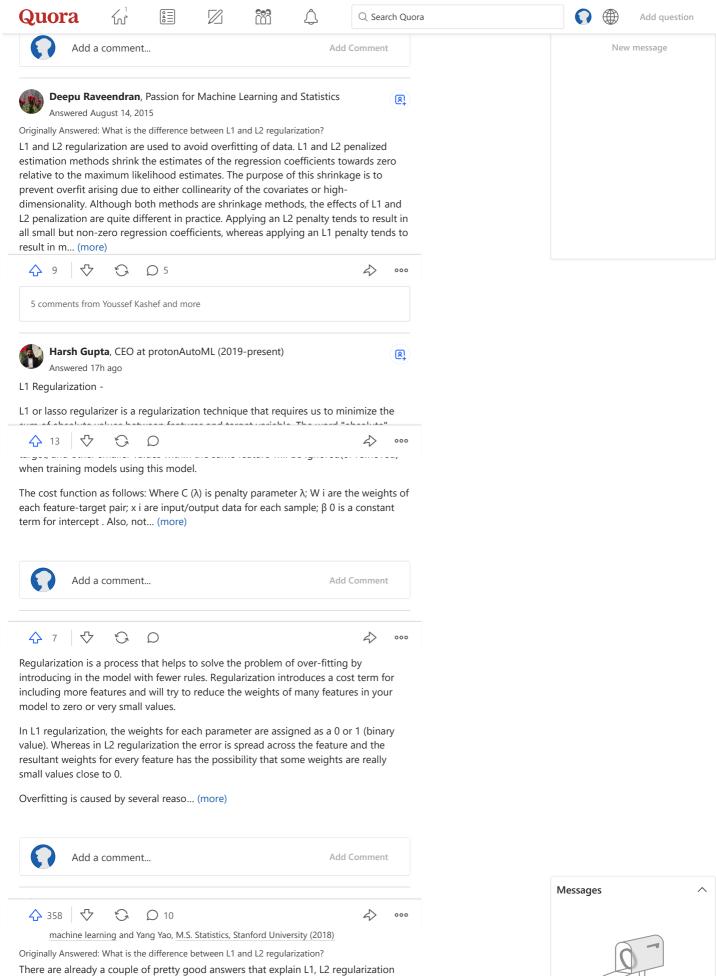
$$\sum_{i=1}^{n} (Y_i - \hat{b}_0 - \hat{b}_1 X_i - \cdots \hat{b} X_p)^2 + \alpha (|\hat{b}_1| + \cdots + |\hat{b}_p|)$$

On the one hand, if we do not apply any penalty (set  $\alpha$  =0), the above formula turns into a regular OLS regression, which may overfit.

On the other hand, the model will probably underfit if we apply a very large penalty (or, a large  $\alpha$  value), because we have falsely penalized all coefficients (the most important ones included).

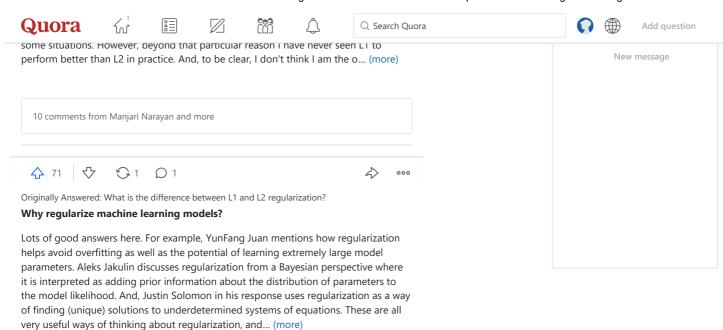
Ridge regression adopts a "squared magnitude" of coefficient tim... (more)





from a theoretical perspective. In my answer I am going to try to complement them from a purely practical point of view. And, in doing so I am going to somewhat

disagree with Andrew Ng.



4>

1 comment from Jo Helmuth



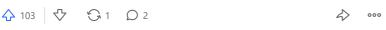
Machine Learning, University College London (2018) and Carl Shan, data scientist, 2014 Eric Schmidt Data Science Fellow

Originally Answered: What is the difference between L1 and L2 regularization?

There are many ways to understand the need for and approaches to regularization. I won't attempt to summarize the ideas here, but you should explore statistics or machine learning literature to get a high-level view. In particular, you can view regularization as a prior on the distribution from which your data is drawn (most famously Gaussian for least-squares), as a way to punish high values in regression coefficients, and so on. I prefer a more naive but somewhat more understandable (for me!) viewpoint.

Let's say you wish to solve the linear problem . Here, is a matrix and is a ve... (more)

12 comments from Manjari Narayan and more



PhD stats and Don van der Drift, In PhD Physics program for 2.5 years at Technische Universiteit Eindhoven, with a focus on statistics and dat...

Originally Answered: What is the difference between L1 and L2 regularization?

The main difference between L1 and L2 regularization is that L1 can yield sparse models while L2 doesn't. Sparse model is a great property to have when dealing with high-dimensional data, for at least 2 reasons.

- Model compression: increasingly important due to the mobile growth
- Feature selection: it helps to know which features are important and which features are not or redundant.

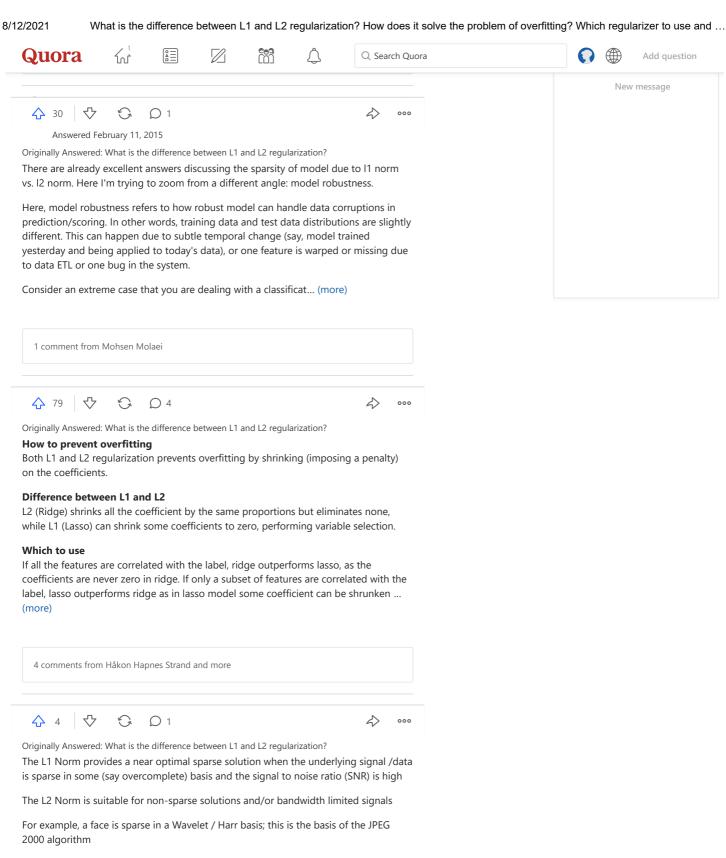
But why does L1 yield sparse models? Despite earlier attempts [1], here's a another simple way to look at it [2].

For simplicity, let's just consider the 1-dimensional case.

-regularized loss function

... (more)





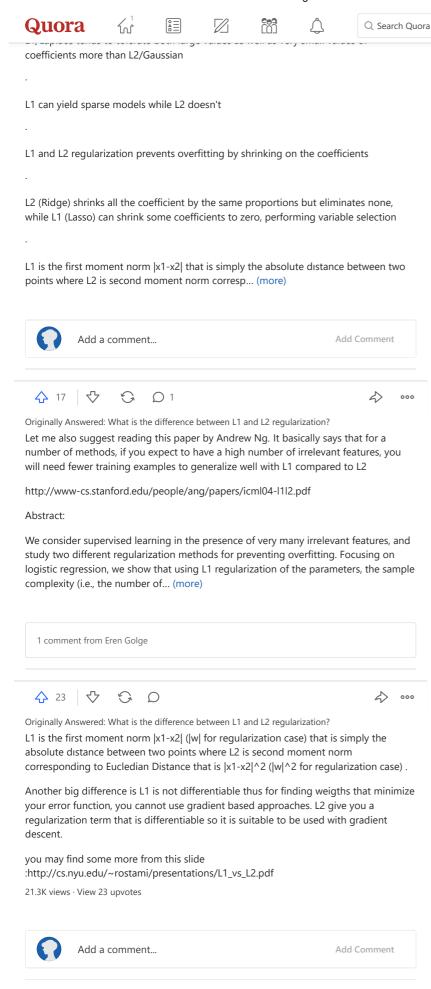
A document, however, is generally not that sparse in the bag of words representation, and L2 methods can work very well here

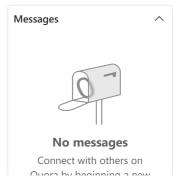
Notice I did not say anything about the noise distribution. Also, the optimal sparse solution is the L0 norm, not the L1 no... (more)



The difference between L1 and L2 regularization are as follows:







Add question

New message

8/12/2021



√1









Q Search Quora





Add question

New message

