

# A Complete understanding of LASSO Regression

---

 [mygreatlearning.com/blog/understanding-of-lasso-regression](https://mygreatlearning.com/blog/understanding-of-lasso-regression)

Great Learning Team

September 4, 2020

Contributed by: **Dinesh Kumar**

## Introduction

---

In this blog, we will see the techniques used to overcome overfitting for a regression model. Regularization is one of the methods widely used to make your model more generalized.

## Lasso Meaning

---

The word “LASSO” stands for **L**east **A**bsolute **S**hrinkage and **S**election **O**perator. It is a statistical formula for the regularisation of data models and feature selection.

## Regularization

---

Regularization is an important concept that is used to avoid overfitting of the data, especially when the trained and test data are much varying.

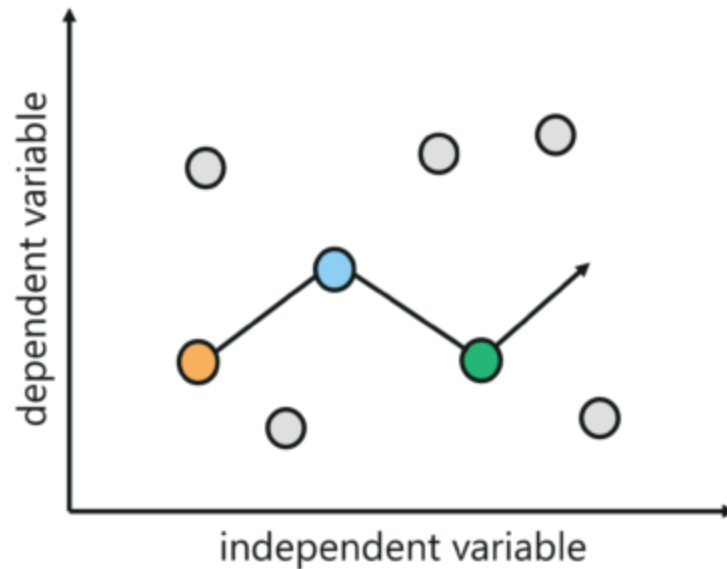
Regularization is implemented by adding a “penalty” term to the best fit derived from the trained data, to achieve a *lesser variance* with the tested data and also restricts the influence of predictor variables over the output variable by compressing their coefficients.

In regularization, what we do is normally we keep the same number of features but reduce the magnitude of the coefficients. We can reduce the magnitude of the coefficients by using different types of regression techniques which uses regularization to overcome this problem. So, let us discuss them.

## Lasso Regularization Techniques

---

There are two main regularization techniques, namely Ridge Regression and Lasso Regression. They both differ in the way they assign a penalty to the coefficients. In this blog, we will try to understand more about Lasso Regularization technique.



## What is Lasso Regression?

---

Lasso regression is a regularization technique. It is used over regression methods for a more accurate prediction. This model uses shrinkage. Shrinkage is where data values are shrunk towards a central point as the mean. The lasso procedure encourages simple, sparse models (i.e. models with fewer parameters). This particular type of regression is well-suited for models showing high levels of multicollinearity or when you want to automate certain parts of model selection, like variable selection/parameter elimination.

Lasso Regression uses L1 regularization technique (will be discussed later in this article). It is used when we have more number of features because it automatically performs feature selection.

## L1 Regularization

---

If a regression model uses the L1 Regularization technique, then it is called Lasso Regression. If it used the L2 regularization technique, it's called Ridge Regression. We will study more about these in the later sections.

L1 regularization adds a penalty that is equal to the absolute value of the magnitude of the coefficient. This regularization type can result in sparse models with few coefficients. Some coefficients might become zero and get eliminated from the model. Larger penalties result in coefficient values that are closer to zero (ideal for producing simpler models). On the other hand, L2 regularization does not result in any elimination of sparse models or coefficients. Thus, Lasso Regression is easier to interpret as compared to the Ridge.

*Also Read: [Python Tutorial for Beginners](#)*

## Mathematical equation of Lasso Regression

---

## Residual Sum of Squares + $\lambda$ \* (Sum of the absolute value of the magnitude of coefficients)

Where,

- $\lambda$  denotes the amount of shrinkage.
- $\lambda = 0$  implies all features are considered and it is equivalent to the linear regression where only the residual sum of squares is considered to build a predictive model
- $\lambda = \infty$  implies no feature is considered i.e, as  $\lambda$  closes to infinity it eliminates more and more features
- The bias increases with increase in  $\lambda$
- variance increases with decrease in  $\lambda$

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

## Lasso Regression in Python

---

For this example code, we will consider a dataset from Machine hack's [Predicting Restaurant Food Cost Hackathon](#).

### About the Data Set

---

The task here is about predicting the average price for a meal. The data consists of the following features.

Size of training set: 12,690 records

Size of test set: 4,231 records

### Columns/Features

---

**TITLE:** The feature of the restaurant which can help identify what and for whom it is suitable for.

**RESTAURANT\_ID:** A unique ID for each restaurant.

**CUISINES:** The variety of cuisines that the restaurant offers.

**TIME:** The open hours of the restaurant.

**CITY:** The city in which the restaurant is located.

**LOCALITY:** The locality of the restaurant.

**RATING:** The average rating of the restaurant by customers.

**VOTES:** The overall votes received by the restaurant.

**COST:** The average cost of a two-person meal.

After completing all the steps till Feature Scaling (Excluding), we can proceed to building a Lasso regression. We are avoiding feature scaling as the lasso regression comes with a parameter that allows us to normalise the data while fitting it to the model.

*Also Read: [Top Machine Learning Interview Questions](#)*

## Lasso regression example

---

```
import numpy as np
```

### Creating a New Train and Validation Datasets

```
from sklearn.model_selection import train_test_split
data_train, data_val = train_test_split(new_data_train, test_size = 0.2, random_state = 2)
```

### Classifying Predictors and Target

```
#Classifying Independent and Dependent Features
#_____
#Dependent Variable
Y_train = data_train.iloc[:, -1].values
#Independent Variables
X_train = data_train.iloc[:, 0 : -1].values
#Independent Variables for Test Set
X_test = data_val.iloc[:, 0 : -1].values
```

### Evaluating The Model With RMLSE

```
def score(y_pred, y_true):
    error = np.square(np.log10(y_pred + 1) - np.log10(y_true + 1)).mean() ** 0.5
    score = 1 - error
    return score
actual_cost = list(data_val['COST'])
actual_cost = np.asarray(actual_cost)
```

### Building the Lasso Regressor

## #Lasso Regression

```
from sklearn.linear_model import Lasso
#Initializing the Lasso Regressor with Normalization Factor as True
lasso_reg = Lasso(normalize=True)
#Fitting the Training data to the Lasso regressor
lasso_reg.fit(X_train,Y_train)
#Predicting for X_test
y_pred_lass =lasso_reg.predict(X_test)
#Printing the Score with RMLSE
print("\n\nLasso SCORE : ", score(y_pred_lass, actual_cost))
```

## Output

**0.7335508027883148**

**The Lasso Regression attained an accuracy of 73% with the given Dataset.**

*Also Read: [What is Linear Regression in Machine Learning?](#)*

## Lasso Regression in R

Let us take “The Big Mart Sales” dataset we have product-wise Sales for Multiple outlets of a chain.

In the dataset, we can see characteristics of the sold item (fat content, visibility, type, price) and some characteristics of the outlet (year of establishment, size, location, type) and the number of the items sold for that particular item. Let’s see if we can predict sales using these features.

Let’s us take a snapshot of the dataset:

Item_Weight	Item_Fat_Content	Item_Visibility	Item_MRP	Outlet_Establishment_Year	Item_Outlet_Sales	Outlet_Size_Medium	Outlet_Size_Small	Outlet_Location_Type_Tier 2
9.3	0	-4.132214606	249.8092	14	3735.138	1	0	0
5.92	1	-3.948779525	48.2692	4	443.4228	1	0	0
17.5	0	-4.088755709	141.618	14	2097.27	1	0	0
19.2	1	-2.716102099	182.095	15	732.38	0	1	0
8.93	0	-2.716102099	53.8614	26	994.7052	0	0	0
10.395	1	-2.716102099	51.4008	4	556.6088	1	0	0
13.65	1	-4.362923154	57.6588	26	343.5528	0	0	0
12.85764518	0	-2.059875358	107.7622	28	4022.7636	1	0	0

## Let’s Code!



## Output



## Ridge and Lasso Regression

---

Lasso Regression is different from ridge regression as it uses absolute coefficient values for normalization.

As loss function only considers absolute coefficients (weights), the optimization algorithm will penalize high coefficients. This is known as the L1 norm.

In the above image we can see, Constraint functions (blue area); left one is for lasso whereas the right one is for the ridge, along with contours (green eclipse) for loss function i.e, RSS.

In the above case, for both regression techniques, the coefficient estimates are given by the first point at which contours (an eclipse) contacts the constraint (circle or diamond) region.

On the other hand, the lasso constraint, because of diamond shape, has corners at each of the axes hence the eclipse will often intersect at each of the axes. Due to that, at least one of the coefficients will equal zero.

However, lasso regression, when  $\alpha$  is sufficiently large, will shrink some of the coefficients estimates to 0. That's the reason lasso provides sparse solutions.

The main problem with lasso regression is when we have correlated variables, it retains only one variable and sets other correlated variables to zero. That will possibly lead to some loss of information resulting in lower accuracy in our model.

That was Lasso Regularization technique, and I hope now you can comprehend it in a better way. You can use this to improve the accuracy of your machine learning models.

## Interpretations and Generalizations

---

### Interpretations:

1. Geometric Interpretations
2. Bayesian Interpretations
3. Convex relaxation Interpretations
4. Making  $\lambda$  easier to interpret with an accuracy-simplicity tradeoff

### Generalizations

1. Elastic Net
2. Group Lasso
3. Fused Lasso
4. Adaptive Lasso

5. Prior Lasso
6. Quasi-norms and bridge regression

## FAQs Related to Lasso Regression

---

### **What is Lasso regression used for?**

**A:** Lasso regression is used for eliminating automated variables and the selection of features.

### **What is lasso and ridge regression?**

**A:** Lasso regression makes coefficients to absolute zero; while ridge regression is a model turning method that is used for analyzing data suffering from multicollinearity

### **What is Lasso Regression in machine learning?**

**A:** Lasso regression is a regularization technique used for more accurate prediction.

### **Why does Lasso shrink zero?**

**A:** The L1 regularization performed by Lasso, causes the regression coefficient of the less contributing variable to shrink to zero or near zero.

### **Is lasso better than Ridge?**

**A:** Lasso is considered to be better than ridge as it selects only some features and decreases the coefficients of others to zero.

### **How does Lasso regression work?**

**A:** Lasso regression uses shrinkage, where the data values are shrunk towards a central point such as the mean value.

### **What is the Lasso penalty?**

**A:** The Lasso penalty shrinks or reduces the coefficient value towards zero. The less contributing variable is therefore allowed to have a zero or near-zero coefficient.

### **Is lasso L1 or L2?**

**A:** A regression model using the L1 regularization technique is called Lasso Regression, while a model using L2 is called Ridge Regression. The difference between these two is the term penalty.

### **Is lasso supervised or unsupervised?**

**A:** Lasso is a supervised regularization method used in machine learning.

While there are ample resources available online to help you understand the subject, there's nothing quite like a certificate. Check out Great Learning's **PG program in Artificial Intelligence and Machine Learning** to upskill in the domain. This course will help you learn from a top-ranking global school to build job-ready AIML skills. This 12-month program offers a hands-on learning experience with top faculty and mentors. On completion, you will receive a Certificate from The University of Texas at Austin, and Great Lakes Executive Learning.

If you are a beginner in the field, taking up a **Python for Machine Learning Free Course** offered by Great Learning Academy can be a great place to start. With the help of this course, you can learn all the basic concepts required to enter the field of Machine Learning and build a career in the same. You will receive a course completion certificate which can be added to your social profiles.

6

**LEAVE A REPLY**

---

**3 × 3 =**