**Regularization in Machine Learning**

Regularization plays an important role in **Machine Learning**. It is a technique to prevent the machine learning model from overfitting by adding extra term or information to it. It mainly regularizes or reduces the **coefficient** of features toward zero but not zero. In simple words, in regularization technique we reduce the magnitude of the features by keeping the same number of features.

**Types of regularizations**

There are mainly **two types** of regularizations techniques, which are given below:

* **Ridge regression or L2 regularization.**
* **Lasso Regression or L1 regularization.**

**Ridge Regression:** Ridge regression is one of the types of linear regression in which a small amount of bias is introduced so that we can get better long-term predictions. Ridge regression is a regularization technique, which is used to reduce the **complexity** of the model. In this technique, the cost function is altered by adding the penalty term to it. The amount of bias added to the model is called Ridge Regression penalty. We can calculate it by multiplying with the lambda to the squared weight of each individual feature.

**Mathematical Formulation (Direct formula):**

**For simple Linear Regression:**

We know the cost function of Linear regression:

**After add Extra term and apply partial derivatives:**

**Now, we can easily see if we change the lambda value then it will impact on ‘m’. if increase lambda value, then ‘m’ value will be decrease.**

**For multiple Linear Regression:**

**We know loss function and I add the extra term:**

**In multiple linear regression we found this thing:**

**Now we can apply partial derivatives and write:**

**Matrix from we can write like this :**

**We apply matrix derivatives, explore this pdf:** <https://github.com/alaminbhuyan/ML_DL_AI_Files/blob/master/Machine%20Learning/Matrix_derivatives_cribsheet.pdf>

**For Gradient Descent:**

**Formula is:**

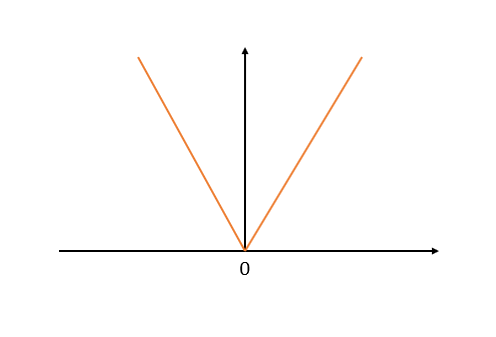
Here, is

**Note:** Lambda value could be zero to infinite. In ridge regression if we increase lambda vlaue over time the coefficients value will be decrease and it going towards zero but never be zero.

**For lasso Regression:**

**After add Extra term and apply partial derivatives:**

**Here, we can’t differentiation mod of ‘m’ or |m|. Because the differentiation of mod of ‘m’ or |m| is not possible at the point of zero.**



**So, we convert it into cases.**

* **If m>0 then |m|=m:**

Here, we multiply by 2 for mathematical convenience.

* **If m<0 then |m|=-m:**
* **If m=0 then:**

**Note:** We know, Lambda value could be zero to infinite. In lasso regression if we increase lambda vlaue over time the coefficients value will be decrease and it going towards zero and at a point it will be zero. That’s why lasso regression also use for features selection because lasso regression makes zero some co-efficient value if we increase lambda vlaue. So, some features are reduced and remaining important features are exits. Lasso is normally used for higher dimension data.

**Elastic Net:**

Elastic Net is the **combination** of Lasso and Ridge regression.