**Regularization** also called smoothing.

It tries to tune the parameters in the target function hoping to make the model less sensitive to fluctuations.

It just adds an extra term to the cost function that is being used to evaluate the model.

Function-

https://miro.medium.com/v2/resize:fit:674/1*3liBCf9xQ8QySEb-8Z3Lbw.png

Where:

* **θ’**s are the factors/weights being tuned.
* **‘λ’** is the regularization rate and it controls the amount of regularization applied to the model. It’s selected using cross-validation.
* **‘R’** is the regularization function and it differs according to the regularization type.
* The model tries to minimize **J(θ)**.

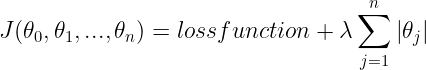
In practice, regularization often leads to a slightly higher bias but significantly reduces the variance. This is what we call the **bias-variance tradeoff**.

 Types of regularization-

# (1) L1 Regularization

*It’s also known as****“L1-Norm”****or****“Lasso Regression”****. Lasso stands for “*Least Absolute Shrinkage and Selection Operator”.

Lasso regression adds a factor of the sum of the absolute values of the model coefficients. Lasso regression tries to minimize the following function:

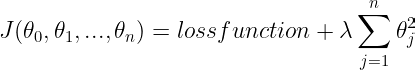


Lasso Regression can also be used for feature selection because the coeﬃcients of less important features are reduced to zero.

# (2) L2 Regularization

It’s also known as**“L2-Norm”** or **“Ridge Regression”**

Ridge regression adds a factor of the sum of the squared values of the model coefficients. Ridge regression tries to minimize the following function:



L2 regularization forces the weights to be small but does not make them zero.

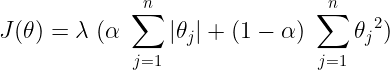
# (3) Elastic Net Regularization

Elastic Net was created to improve the Lasso regression whose variable selection process can be too dependent on the data and thus unstable.

**Lasso Regression has some limitations:**

* If the number of variables/features (n) is greater than the number of training samples, Lasso selects at most ‘n’ variables. So the number of selected features/variables is bounded by the number of samples.
* If some features are highly corelated, Lasso tends to select one feature from them and ignore the others. This may result in an in-accurate model, as those features may be important.

To solve that, Elastic Net combines the properties of both Ridge and Lasso and it tries to minimize the following loss function:



# When to use which?

* Ridge Regression is suitable if you want to keep all the features and avoid that the model becomes over sensitive to the noise/fluctuations in the training data.
* If you think that only few features are useful, then it would be better to use Lasso Regression or Elastic Net as they sets the weights of less-important features to zero.