Faculty of Engineering
Cairo University

Cairo University

Faculty of Engineering

Systems and Biomedical Department

Linear Regression Implementation

Submitted by:

Ashar Seif Al-Naser Saleh

Sec: 1 BN: 9

SBE452_AI

Dr.Inas A. Yassine

6 November, 2021

What happens if we use lasso regression?

Lasso Regression is also another linear model derived from Linear Regression which shares the same hypothetical function for prediction.

Linear Regression model considers all the features equally relevant for prediction. When there are many features in the dataset and even some of them are not relevant for the predictive model. This makes the model more complex with a too inaccurate prediction on the test set (or overfitting). Such a model with high variance does not generalize on the new data. So, Lasso Regression comes for the rescue. It introduced an L1 penalty (or equal to the absolute value of the magnitude of weights) in the cost function of Linear Regression. The modified cost function for Lasso Regression is given below.

$$\frac{1}{m}\left[\sum_{i=1}^{m}\left(y^{(i)}-h\left(x^{(i)}\right)\right)^{2}+\lambda\sum_{j=1}^{n}w_{j}\right]$$

Mathematical Intuition:

During gradient descent optimization, added L1 penalty shrunk weights close to zero or zero. Those weights which are shrunken to zero eliminates the features present in the hypothetical function. Due to this, irrelevant features don't participate in the predictive model. This penalization of weights makes the hypothesis simpler which encourages the sparsity (model with few parameters).

If the intercept is added, it remains unchanged.

We can control the strength of regularization by hyper parameter lambda. All weights are reduced by the same factor lambda.

Different cases for tuning values of lambda.

1. If lambda is set to be 0, Lasso Regression equals Linear Regression.
2. If lambda is set to be infinity, all weights are shrunk to zero.

If we increase lambda, bias increases if we decrease the lambda variance increase. As lambda increases, more and more weights are shrunk to zero and eliminates features from the model.