RESULT ANALYSIS

Regularization is specifically used in machine learning to avoid over-fitting issues faced by the trained model and to reduce the loss function(MSE). There are 2 types of regularizations(L1 & L2). In the L1 regularization which is also called as Lasso Regularization, if the weight(w) is greater 0 than 0 then it is applied as "**1*(learning rate)*w**" and otherwise "**-1*(learning rate)*w**". In the L2 regularization, also called as Ridge regularization, it is applied as "**2*(learning rate)*w**".

When we performed regularization on the red wine dataset, we found that L1 regularization achieved an MSE of "**0.4199**" using batch gradient descent and "**0.2805**" with mini-batch gradient descent which proves its effectiveness by significantly reducing the loss function. In this particular scenario, L2 regularization has been as effective as L1 maybe due to the dataset's characteristics or variations in the features. The values of L1 regularization are better for both Gradient Descent and Batch Gradient Descent.
For the white wine dataset, for the batch gradient descent, it is evident that the MSE is slightly higher in L1(**0.5912**) than without any regularization(**0.5619**). Hence, it has not provided the best MSE but yet it is better than the L2 with a significant difference between the two with MSE obtained from L2 is "**0.8542**". In the case of Mini-Batch Gradient Descent, L1 has significantly reduced the MSE with a value "**0.4736**" wherein it is extremely without regularization which produced an MSE of "**1.183**". L2 produced an MSE of "**0.5030**" which is still not preferable to L1 but yet is way better than done without any regularization. This comparatively shows that L1 regularization is more effective in this case, especially in mini-batch gradient descent, where it produced the lowest MSE. We can also see how important regularization is in the case of Mini-Batch Gradient Descent.

In conclusion, L1 regularization proved more effective, especially in mini-batch gradient descent, where it consistently produced lower MSE than L2 and no regularization. Regularization was crucial in reducing overfitting, particularly in mini-batch gradient descents of both the datasets.