**Analysis of Regularization Approaches**

**Results can be accessed at results\_summary.json (../model\_outputs)**

{

"plain": {

"train\_mse": 0.42106045980473755,

"valid\_mse": 0.40494552844023796,

"test\_mse": 0.4167108229543469

},

"minibatch\_plain": {

"train\_mse": 0.4207878089449882,

"valid\_mse": 0.4063107120981717,

"test\_mse": 0.41609454746073654

},

"l2": {

"alpha": 0.1,

"train\_mse": 0.42269423804638884,

"valid\_mse": 0.4043037010254559,

"test\_mse": 0.41979826247729335

},

"plain\_after\_l2\_drop": {

"dropped\_feature": "citric acid",

"train\_mse": 0.42176729500684357,

"valid\_mse": 0.4037994420706724,

"test\_mse": 0.41767321770767224

},

"l1": {

"alpha": 0.0,

"train\_mse": 0.42106045980473755,

"valid\_mse": 0.40494552844023796,

"test\_mse": 0.4167108229543469

},

"plain\_after\_l1\_drop": {

"dropped\_feature": [

"residual sugar"

],

"train\_mse": 0.42118388479068164,

"valid\_mse": 0.4065197575817783,

"test\_mse": 0.41744250687729084

}

}

**Plots**

**1)Batch plain**

**A diagram of a line graph

Description automatically generated with medium confidence**

**2)After L1 drop**

**A graph of a drop of alcohol

Description automatically generated**

**3)After L2 drop**

**A graph of a drop of alcohol

Description automatically generated with medium confidence**

**4)L1 – alpha – 0.0**

**A graph of a line

Description automatically generated with medium confidence**

**5)L2 alpha – 0.1**

**A graph of a line

Description automatically generated with medium confidence**

**Base Model:**

Train MSE: 0.421

Validation MSE: 0.405

Test MSE: 0.417

They are consistent across different splits which means the model generalized reasonably with little to no overfitting. The optimization choice did not significantly affect performance for this dataset size.

**L2 Regularization**

L2 regularization penalizes the squared magnitude of weights and then shrinks them towards zero while keeping all feature. With best penalty strength (alpha = 0.1):

* Validation error decreased slightly to 0.4043 (vs 0.4049).
* Train error increased slightly because regularization restricts model flexibility.
* Test error rose noticeable enough (0.4198 vs 0.4167).

After dropping the least influential feature (citric acid), validation error improved to 0.4038, which is the best performance. Removing weak features such as citric acid can very slightly improve generalization.

**L1 Regularization**

L1 penalizes the absolute magnitude of weights which means some coefficients are set exactly to zero through automatic feature selection. Alpha = 0.0 was the best for this model, by defaulting to the plain model. L1 doesn’t affect much or provide improvement for this dataset. Even after dropping a feature (residual sugar) , it actually worsened.

**Conclusion**

L2 seems to be the most effective among the given models, while L1 showed no improvement and maybe even worsening. Dataset was not prone to overfitting considerably.