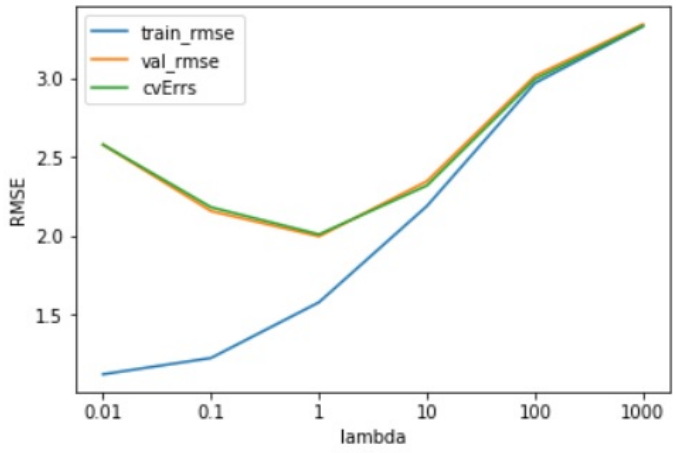
Q. 3.2

1. Solve Ridge to predict the number of points a Wine will receive. Run Ridge on the training set, with lambda = 0.01, 0.1, 1, 10, 100, 1000. At each solution, record the root-mean-squared-error (RMSE) on training, validation and leave-one-out-cross-validation data.

****
lambda-> 0.01
rmse of cvErrs= 2.580103521790279
train_rmse 1.1205188247625724
val_rmse 2.5791868257078514
****
lambda-> 0.1
rmse of cvErrs= 2.182106499121022
train_rmse 1.2238119656960245
val rmse 2.1574812686463836
****
lambda-> 1
rmse of cvErrs= 2.009474060889735
train_rmse 1.5780360753182243
val rmse 1.9967715113690583
****
lambda-> 10
rmse of cvErrs= 2.3201351657980895
train_rmse 2.189953395751638
val_rmse 2.347705513916286
****
lambda-> 100
rmse of cvErrs= 2.9965727187353157
train_rmse 2.9709419656108045
val_rmse 3.017105238664671
****
lambda-> 1000
rmse of cvErrs= 3.3353399803872232
train_rmse 3.3316125921530397
val_rmse 3.345415090847554

Plot the train, validation and leave-one-out-cross-validation RMSE values together on a plot against Lambda. Label each curve in the plot.



2. What lambda achieve the best LOOCV performance? For the model using this lambda, report the objective value, the sum of square errors (on training data), the value of the regularization term.

Lambda having best LOOCV performance is 1
Objective Value: 17200.94056872298
Sum of square errors: 12450.989275028724
Regularization Value 4749.951293694259

3. Using the model that you computed using lambda that achieves best LOOCV performance, list the top 10 most important features and the top 10 least important features. Comment if the weights make sense intuitively.

Top 10 most imp features:
infused
pineapple orange
red
flavors nice
sweet black
little heavy
new french
future
currant cola
cocktail

Top 10 least imp features:
offers
light body
highlights
franc petit verdot
framed
tannins frame
tannins finish
sour
flavors black cherry
Oakville

Comment: It seems features of wine flavours are contributing more to the predictions.

4. Use your model to predict the points for the reviews in test data. Report the RMSE.

RMSE: 1.91955