BerkeleyX: CS190.1x Scalable Machine Learning

LAB3: MILLIONSONG REGRESSION PIPELINE (100/100 points)

Once you have successfully run the test notebook, you can submit to the course autograder by first **exporting lab3 as a Python (.py) file**, and then using the file chooser to select your file and clicking *"Check"* to submit your code.

Before submitting your assignment, please ensure that your submission follows these guidelines:

- Only use the following libraries: standard python libraries, numpy, pyspark, and test_helper (the autograder library).
- Don't include any extraneous code as the autograder will timeout if a submission takes too long.
- Only change sections of code where you see FILL IN. Changing other parts of the code, including directory paths, may cause the code to fail the autograder's tests.
- The autograder grades submissions using the same tests that are included in your notebook. Hence, please run the local tests before submitting to the autograder.

For further background on the autograder please visit this page with the information from Week 0 and the autograder FAQ. To check the status of your submission, please visit the autograder submission website.

This is the autograder only for your submission of "Lab3 - Linear Regression". Please **DO NOT submit other labs to this autograder.**

Choose Files No file chosen

```
Load and check the data (1a)
----------------------------
All tests passed
Using LabeledPoint (1b)
-----------------------
All tests passed
Find the range (1c)
-------------------
All tests passed
Shift labels (1d)
-----------------
All tests passed
Training, validation, and test sets (1e)
----------------------------------------
All tests passed
Average label (2a)
------------------
All tests passed
Root mean squared error (2b)
----------------------------
All tests passed
```

```
Training, validation and test RMSE (2c)
---------------------------------------
All tests passed
Gradient summand (3a)
---------------------
All tests passed
Use weights to make predictions (3b)
------------------------------------
All tests passed
Gradient descent (3c)
---------------------
All tests passed
Train the model (3d)
--------------------
All tests passed
LinearRegressionWithSGD (4a)
----------------------------
All tests passed
Predict (4b)
------------
All tests passed
Evaluate RMSE (4c)
------------------
All tests passed
Grid search (4d)
----------------
All tests passed
Vary alpha and the number of iterations (4e)
--------------------------------------------
All tests passed
Add two-way interactions (5a)
-----------------------------
All tests passed
Build interaction model (5b)
----------------------------
All tests passed
Evaluate interaction model on test data (5c)
--------------------------------------------
All tests passed
-- 20 cases passed (100.0%) --


Your submission token ID is 2050801-b9c3ea967ccb0be9f07e27de2cb17c93:ip-172-31-27-255
Please include this submission token ID when you need support for your code submission.
Your anonymous student ID is d9e65bc8d252ec579ef766790c87772e. Do not post this ID on
Piazza.
```

CHECK **SAVE** *You have used 1 of 10 submissions*