**Bayesian Linear Regression:**

**Overview:**

Our goals in this assignment are to explore the role of regularization in linear regression and to investigate two methods for model selection (evidence maximization and cross validation). In all our experiments we report the performance in terms of the mean square error:

A picture containing text, watch, clock

Description automatically generated

where the number of examples in the corresponding dataset is N.

**Data:**

We have 4 datasets and each dataset comes in 4 files with the training set in train-name.csv the corresponding labels (regression values) in trainR-name.csv and similarly for test set. We have two real datasets crime and wine and two artificial datasets artsmall and artlarge.

For the artificial data you can compare the MSE results to the MSE of the hidden true functions generating the data that give 0.533 (artsmall), and 0.557 (artlarge).

**Task 1: Regularization**

In this part we use regularized linear regression, i.e., given a dataset, the solution vector w is given by:



For each of the 4 datasets, we plot the training set MSE and the test set MSE as a function of the regularization parameter λ (we use integer values in the range 0 to 150).

The experiments in this task tell us which value of λ is best in every case in hindsight. That is, we need to see the test data and its labels in order to choose λ. This is clearly not a realistic setting and it does not give reliable error estimates. The next two tasks investigate methods for choosing λ automatically without using the test set.

**Task 2: Model Selection using Cross Validation**

In this part we use 10 fold cross validation on the training set to pick the value of λ in the same range as above, then retrain on the entire train set and evaluate on the test set.

We implement this scheme, apply it to the 4 datasets and report the values of λ selected, associated MSE and the run time.

**Task 3: Bayesian Model Selection**

In this part we consider the formulation of Bayesian linear regression with the simple prior w given by:

w ∼ N (0, 1/α \* I)

The evidence function (and evidence approximation) gives us a method to pick the parameters α and β.

The solution is given by equations below:

A picture containing watch

Description automatically generated

Text

Description automatically generated

A picture containing text, watch

Description automatically generated

where mN and SN are given by:

Text

Description automatically generated with medium confidence

These yield an iterative algorithm for selecting α and β using the training set. We can then calculate the MSE on the test set using the MAP (mN ) for prediction.

This scheme is pretty stable and converges in a reasonable number of iterations. We can initialize α, β to random values in the range [1, 10] and stop the algorithm when the difference in α, β values is < 0.0001.

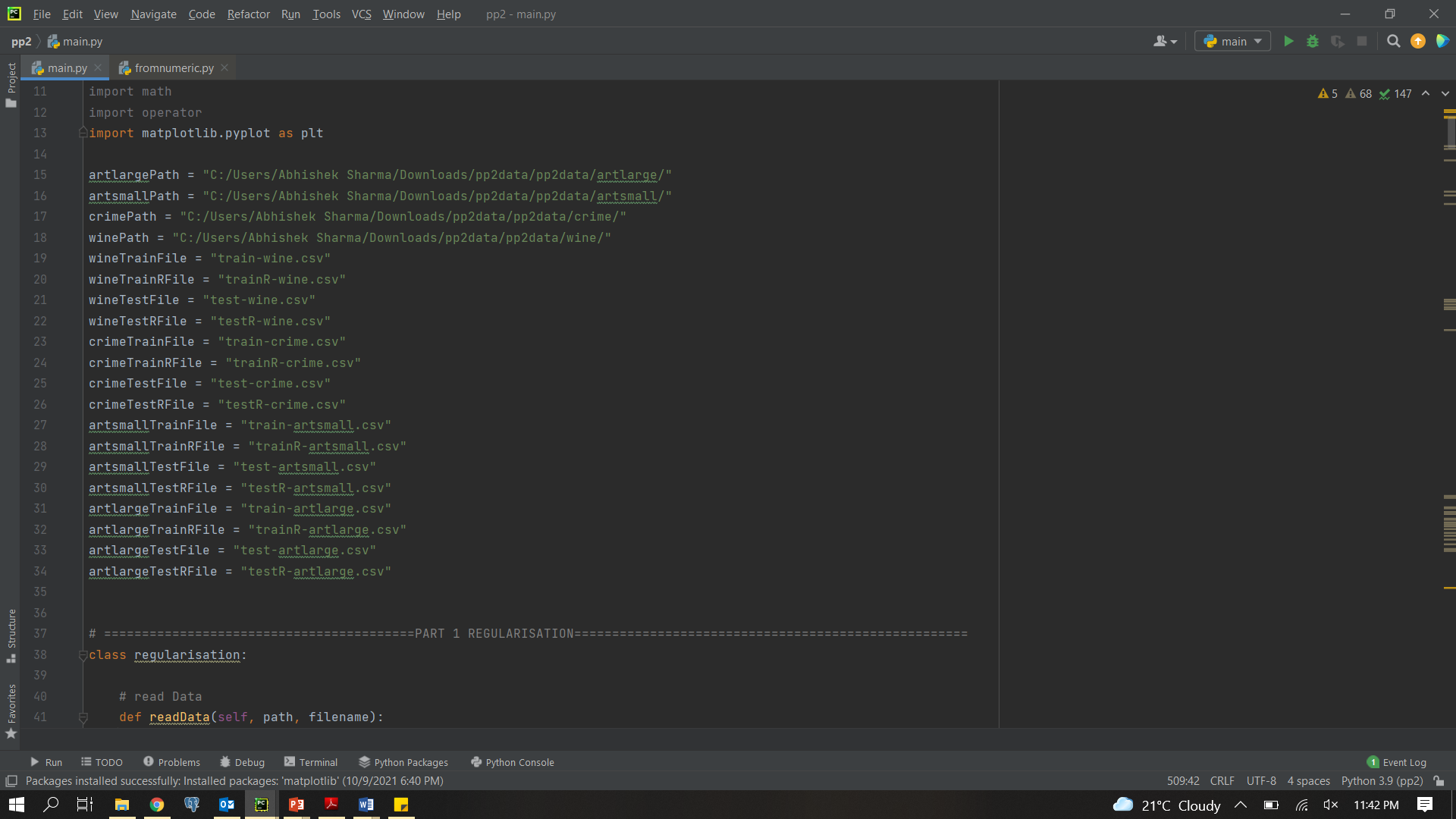
We then implement this scheme, apply it to the 4 datasets and report the values of α, β, the effective λ = α/β, the associated MSE and the run time.

**Info about the code:**

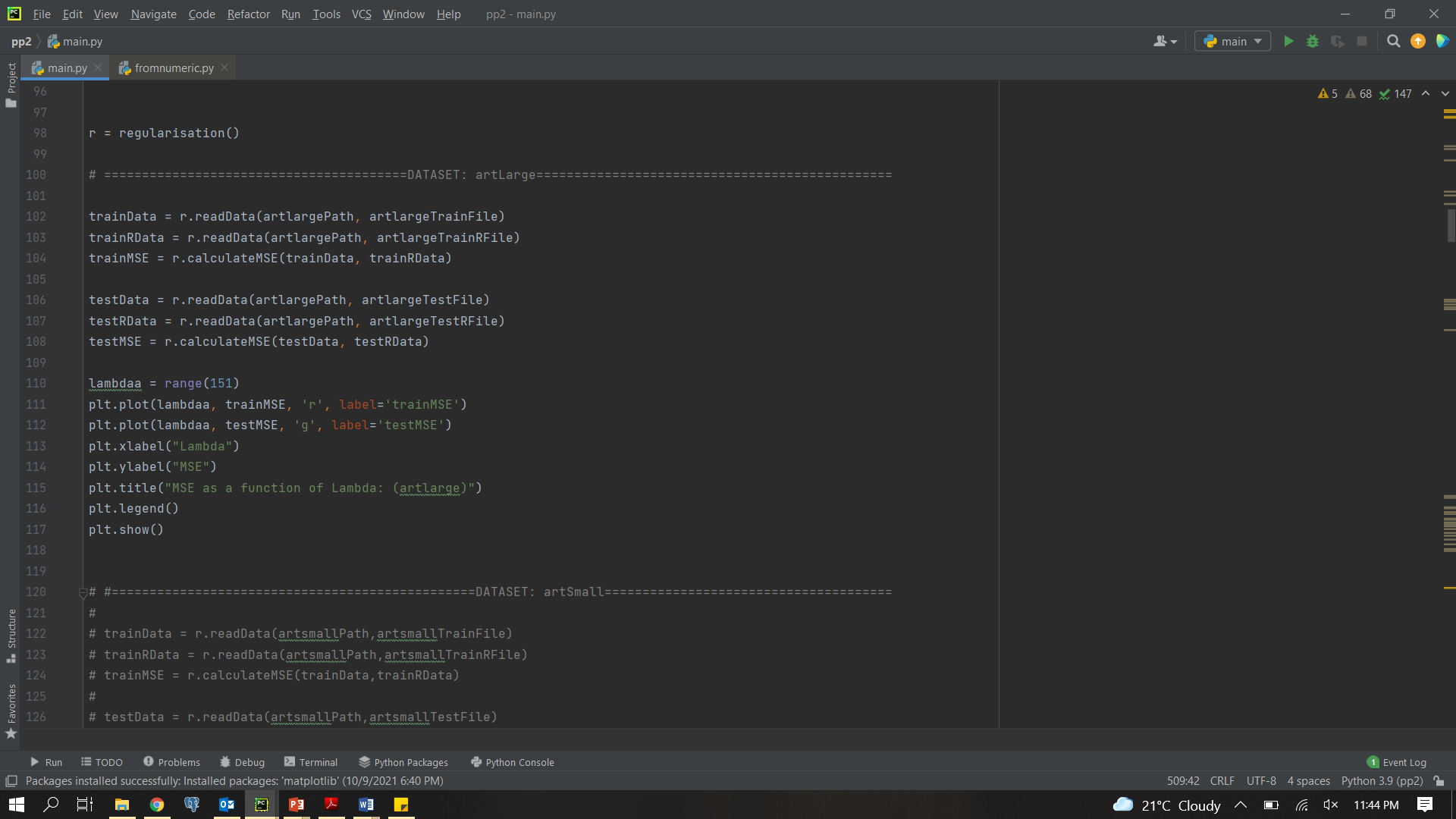
1. The code is divided into three sections: Part 1: Regularisation, Part 2: Cross Fold and part 3: Bayesian Model Selection.
2. In each part, the code is further divided into 4 parts according to each of the 4 data sets.
3. Out of these 4 parts of code, one is left uncommented and the remaining three are commented.

**Steps to run the code:**

1. Change the path values in the path variables according to the new location where the code will be run. The variables are divided into the base path and the respective file names of the datasets.



1. Move to the part which you want to execute and uncomment the code for that particular dataset that you want to test the code on.



1. Run the code.