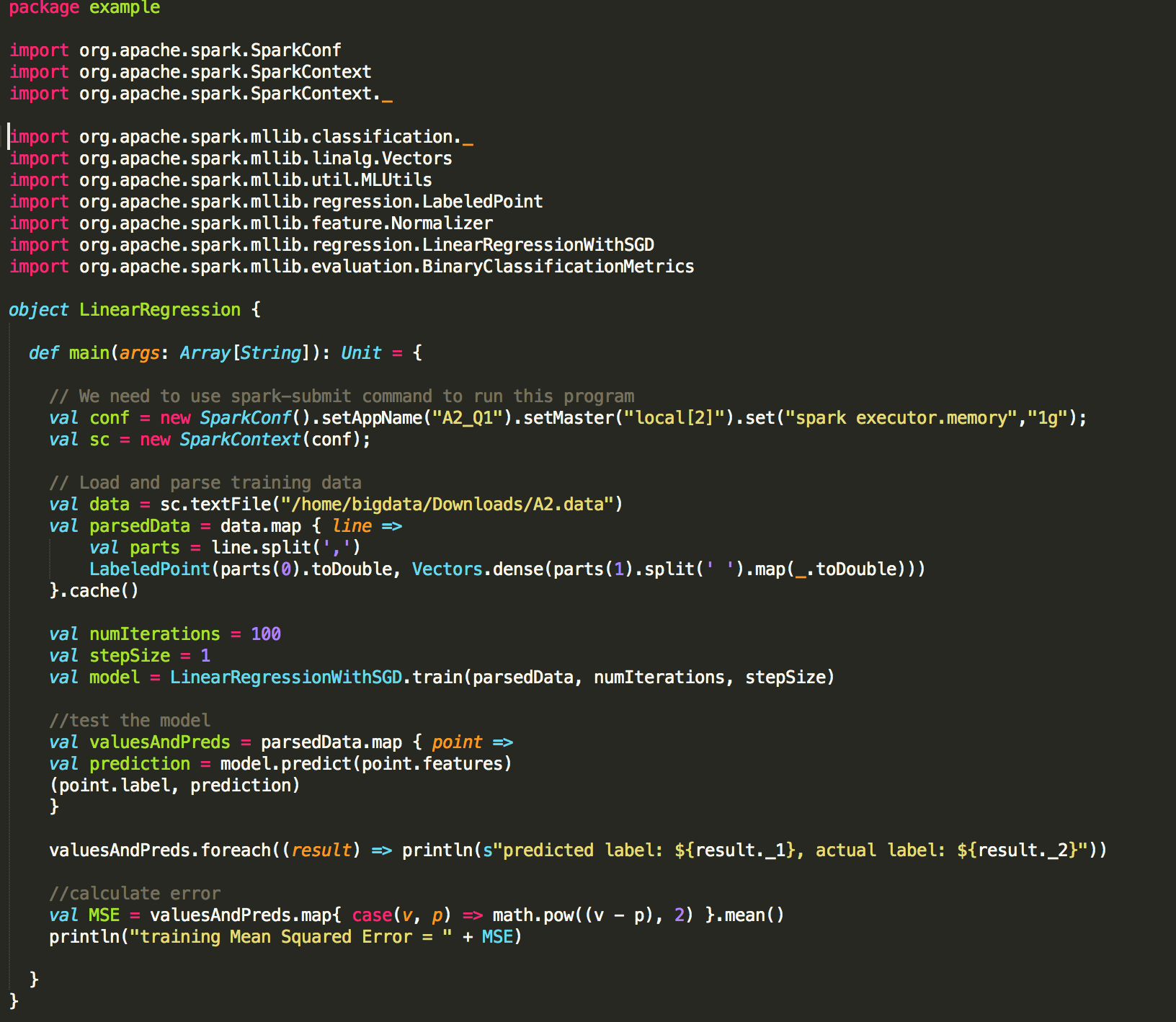
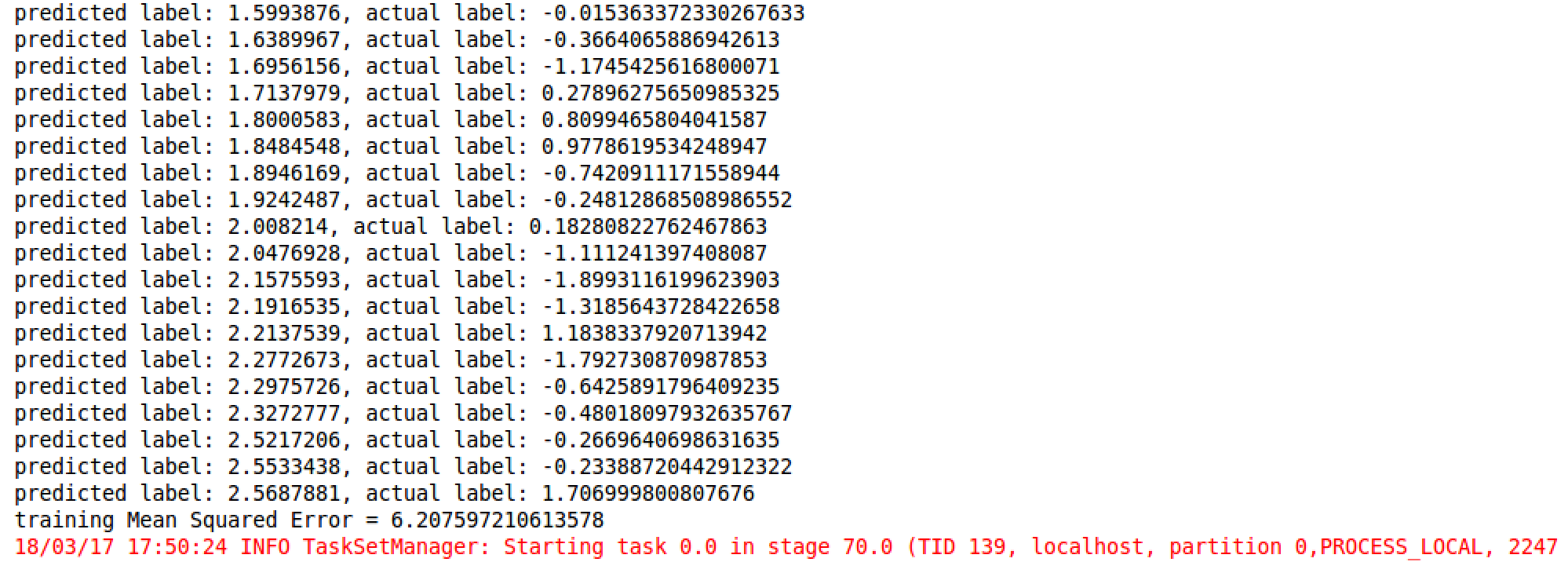
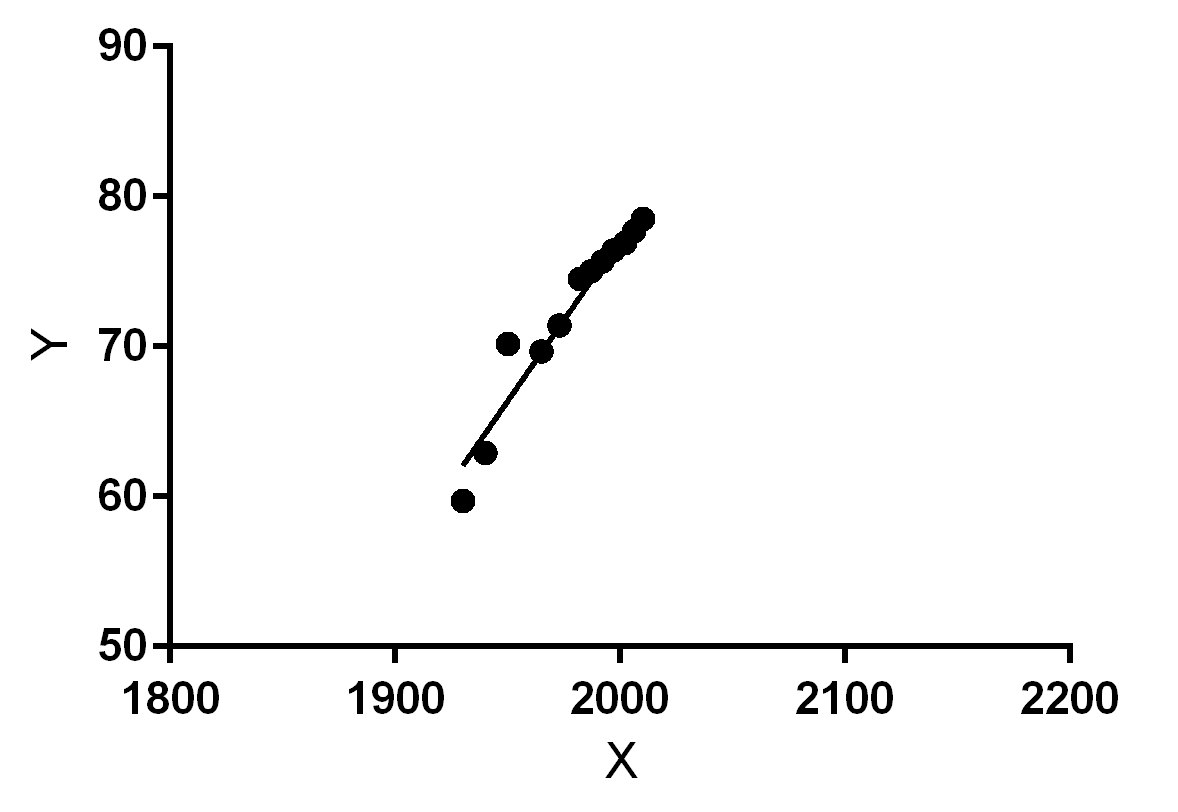
1. Scala program for Linear Regression Model
   1. Code: 
   2. Partial Output:
2. Life expectancy prediction
   1. Based on the whole data, can you predict the life expectancy for individual born in 2015?



By using **linear regression** model, we can predict the life expectancy for individual born in 2015, with a best fitting line , which X is the year of birth and Y is the life expectancy. According to the equation, when X = 2015, Y = 80.3415.

* 1. If life expectancy is 78.7, how to improve your prediction?

From 2.1, we can see that the prediction is 80.3415, while the actual result is 78.7. In this case, all the data are viewed as training set to predict the life expectancy in 2015. However, this might cause an overfitting problem. The problem is that the best fitting line for the training data but might not be the best line to predict new data.

Therefore, to improve the accuracy of the prediction, we can split the dataset into training data and testing data(ex: 75:25) Testing data is used to estimate how well your model has been trained and to estimate the error rate. When we apply the prediction model on our test set, we can have an idea of the performance of the model on unseen data.