CS480 Java and Internet Applications
Week 10

6. Project: Machine Learning
   - Step 1: Understand why a machine can learn
   - Step 2: Study the Machine Learning algorithm Linear Regression.
   - Step 3: Implement Linear Regression.
     - Step 3.1: Collect the data: Read the data from the file "input.text"
       - References
         - Read multi-column data
     - Step 3.2: Create the model: Find "a" and "b" of a "Linear Regression Equation(y) = a + bx" based on the content of "input.txt".
       - The file "input.txt" has these values:

60	3.1
61	3.6
62	3.8
63	4
65	4.1

     - Step 3.3: Prediction: If x=64, predict the y value.
   - Step 4: You are encouraged to create a formal report for this project and post the report on Github.
   - References
     - Study Summary of Three Basic Machine Learning Algorithms
     - Read/Write Primitive Data Types
     - How to open a txt file and read numbers in Java
     - Read CSV file column by column

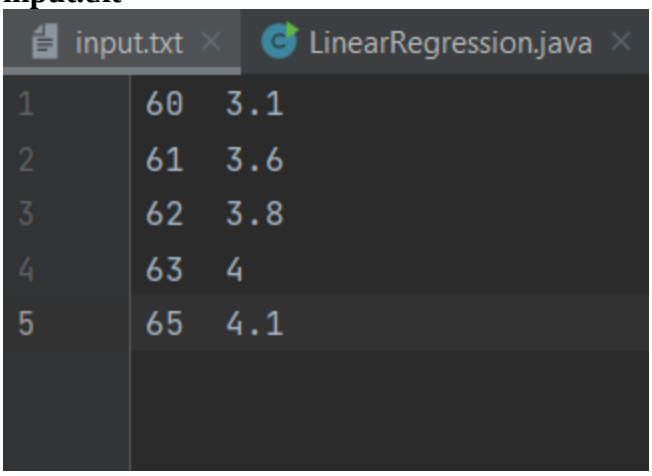
```
import java.io.BufferedReader;
import java.io.FileReader;
import java.io.IOException;
import java.util.ArrayList;
import java.util.List;
import java.util.Scanner;
import java.util.stream.IntStream;

public class LinearRegression {
   private static final String FILE_LOCATION = "src/main/java/input.txt";
   static final List<Float> columnXValues = new ArrayList<>();
   static final List<Float> columnYValues = new ArrayList<>();

   private static void readFile() {
      try {
         BufferedReader bReader = new BufferedReader(new
FileReader(LinearRegression.FILE_LOCATION));
         String line = bReader.readLine();
         float x, y;
         System.out.println("X    Y");
         while (line != null) {
            String[] fields = line.split("\\s+");
            // Get the 1st column
            x = Float.parseFloat(fields[0]);
            // Adding the 1st column values to a list
            columnXValues.add(x);
            // Get the second column
            y = Float.parseFloat(fields[1]);
            // Adding the 2nd column values to a list
            columnYValues.add(y);
            System.out.println(x + " " + y);
            System.out.println();
            // Read next line
            line = bReader.readLine();
         }
      } catch (IOException e) {
         System.out.println("Error reading the file, message: ");
         e.printStackTrace();
      }
   }

   private static void calculateLinearRegression() {
      System.out.println("The values of X(First Score) are:" +
LinearRegression.columnXValues);
      System.out.println("The values of Y(Second Score) are:" +
LinearRegression.columnYValues);
```

```
        float sumX = Util.sum(LinearRegression.columnXValues);
        System.out.println();
        System.out.println("Sum of first scores(ΣX):" + sumX);

        float sumY = Util.sum(LinearRegression.columnYValues);
        System.out.println("Sum of second scores(ΣY):" + sumY);

        List<Float> multiplyXY = Util.mul(LinearRegression.columnXValues,
LinearRegression.columnYValues);
        System.out.println();
        System.out.println("Product of first and Second Scores (XY): " + multiplyXY);

        float sumXY = Util.sum(multiplyXY);
        System.out.println("Sum of the product of first and Second Scores (ΣXY): " + sumXY);
        System.out.println();

        List<Float> multiplyXX = Util.mul(LinearRegression.columnXValues,
LinearRegression.columnXValues);
        System.out.println("Square of First Scores(X*X): " + multiplyXX);

        float sumXX = Util.sum(multiplyXX);
        System.out.println("Sum of squares of First Scores(ΣX*X): " + sumXX);
        System.out.println();

        double b = Util.getB(sumX, sumY, sumXY, sumXX);
        System.out.println("Value of Slope(b) is: " + b);

        double a = Util.getA(sumX, sumY, b);
        System.out.println("Value of Intercept(a) is: " + a);
        System.out.println();
        System.out.println("Regression Equation(y) = a + bx ");
        System.out.println("y=" + a + "+" + b + "x");
        System.out.println("Suppose if we want to know the approximate y value for the variable x
= 64. " +
                "Then we can substitute the value in the above equation.");

        double ypred = Util.getYpred(b, a);
        System.out.println("y predicted for value x=64 is " + ypred);
    }

    public static void main(String[] args) {
        readFile();
        calculateLinearRegression();
    }
```

```
}

final class Util {
   private Util() {
   }

   static float sum(List<Float> list) {
      float sum = 0;
      for (float i : list) {
         sum += i;
      }
      return sum;
   }

   static List<Float> mul(List<Float> list, List<Float> list1) {
      float[] result = new float[list.size()];
      List<Float> mulResult = new ArrayList<>();

      IntStream.range(0, list.size()).forEach(i -> {
         result[i] = list.get(i) * list1.get(i);
         mulResult.add(result[i]);
      });

      return mulResult;
   }

   static double getA(float sumX, float sumY, double b) {
      return Math.round((sumY - b * sumX) / LinearRegression.columnXValues.size() *
1000.00) / 1000.00;
   }

   static double getB(float sumX, float sumY, float sumXY, float sumXX) {
      return Math.round((LinearRegression.columnXValues.size() * sumXY - sumX * sumY) /
            (LinearRegression.columnXValues.size() * sumXX - Math.pow(sumX, 2.0)) * 100.00)
/ 100.00;
   }


   static double getYpred(double b, double a) {
      Scanner sc=new Scanner(System.in);
      System.out.println("Enter the X value for which you want to predict Y");
      float x=sc.nextFloat();
      return Math.round((a + b * x) * 100.00) / 100.00;
   }
}
```

CS480 Java and Internet Applications
Week 10

input.txt



OUTPUT

```
X    Y
60.0 3.1

61.0 3.6

62.0 3.8

63.0 4.0

65.0 4.1

The values of X(First Score) are:[60.0, 61.0, 62.0, 63.0, 65.0]
The values of Y(Second Score) are:[3.1, 3.6, 3.8, 4.0, 4.1]

Sum of first scores(ΣX):311.0
Sum of second scores(ΣY):18.6

Product of first and Second Scores (XY): [186.0, 219.59999, 235.59999, 252.0, 266.5]
Sum of the product of first and Second Scores (ΣXY): 1159.7

Square of First Scores(X*X): [3600.0, 3721.0, 3844.0, 3969.0, 4225.0]
Sum of squares of First Scores(ΣX*X): 19359.0

Value of Slope(b) is: 0.19
Value of Intercept(a) is: -8.098

Regression Equation(y) = a + bx
y=-8.098+0.19x
Suppose if we want to know the approximate y value for the variable x = 64. Then we can substitute the value in the above equation.
Enter the X value for which you want to predict Y
64
y predicted for value x=64 is 4.06
```

CS480 Java and Internet Applications
Week 10

Link to the Google Slides:

https://docs.google.com/presentation/d/1Co64mZRfnMap-6ViKiLPtbFtMt6kSQgD_xX5TbOqVpQ/edit?usp=sharing

Link to GitHub:

https://github.com/HarshineeRoopakula/Java-Programming/tree/main/Machine%20Learning%20Project%20with%20Java