**“An Introduction to Statistical Learning” (ISLR) Chapter 1 & 2 and finish the following questions. Page 63, Chapter 2.4 Exercises:**

2. Explain whether each scenario is a classification or regression problem and indicate whether we are most interested in inference or prediction. Finally, provide n and p.

1. We collect a set of data on the top 500 firms in the US. For each firm we record profit, number of employees, industry, and the CEO salary. We are interested in understanding which factors affect CEO salary.

**Understanding CEO’s Salary:**

* **Type:** Regression problem, because we are interested in understanding the factors that affect the CEO’s salary which is a target variable, and a continuous variable.
* **Interest:** Inference, because we are trying to draw conclusions from the given factors (profit, number of employees, industry) affect the target variable, in this case it’s the CEO’s salary and not trying to predict the salary of the CEO.
* **n (Number of observations) =** Top 500 firms in the US

**p (predictors) =** profit, number of employees, industry.

1. We are considering launching a new product and wish to know whether it will be a success or a failure. We collect data on 20 similar products that were previously launched. For each product we have recorded whether it was a success or failure, price charged for the product, marketing budget, competition price, and ten other variables.

**Product Success Prediction:**

* **Type:** Classification problem, as we are interested in knowing the success and failure rate of the new product (target variable), which is categorical.
* **Interest:** Prediction, because we are interested in knowing whether launching of the new product will be success or a failure based on the collection of the similar products that were previously launched.
* **n =** 20 similar products launched previously.

**p =** price charged for the product, marketing budget, competition price, ten other variables.

1. We are interested in predicting the % change in the USD/Euro exchange rate in relation to the weekly changes in the world stock markets. Hence, we collect weekly data for all of 2012. For each week we record the % change in the USD/Euro, the % change in the US market, the % change in the British market, and the % change in the German market.

**USD/Euro Exchange Rate Prediction:**

* **Type:** Regression problem, as we are interested in knowing the % change in the USD/Euro exchange rate, which is continuous variable.
* **Interest:** As it is mentioned in the question, we are interested in predicting the % change in the USD/Euro exchange rate.
* **n =** Data collected weekly for the entire year 2012.

**p =** the % change in the US market, the % change in the British market, and the % change in the German market.

4. You will now think of some real-life applications for statistical learning.

1. Describe one real-life applications in which classification might be useful. Describe the response, as well as the predictors. Is the goal of each application inference or prediction? Explain your answer.

* **Students Performance Index:** Suppose we take a dataset of 500 students, where each student is categorized by variables such as GPA, IQ level and study hours. The objective of this task to predict whether the students get placement or not after their graduation.

**n** = 500 students

**p** = student’s GPA, IQ level, Study hours.

**Goal:** Prediction. The objective of this task is to predict whether the students will get placed or not.

**Response Variable** = Students’ placement (Yes/No, which is categorical)

1. Describe one real-life application in which regression might be useful. Describe the response, as well as the predictors. Is the goal of each application inference or prediction? Explain your answer.

* **House sale price prediction** is a great example of regression analysis. The price of a house depends on several features.

**Response Variable** = House Sale Price

**Predictors** = Number of bedrooms, square-feet, income, and location

**Goal:** Prediction of the house price based on various factors.

1. Describe one real-life application in where clustering might be useful.

* **Customer Segmentation:** E- commerce business widely uses clustering (unsupervised learning), as a technique to group their customers based on different features.

**Goal:** Grouping or Segmentation and Inference.

**Predictors** = Purchases, Preferences, and demographics.

The target is to identify distinct customers and apply the marketing techniques to provide personalized services.