----------------------------------------------------------------------------

**Exploring the Non-Linear Relationship Between Years of Experience and Salary Using Kernel Regression**

**Introduction**

In today's competitive tech industry, understanding the factors that influence employee salaries is essential for both management and HR practices. One such critical factor is years of experience, which generally has a positive correlation with salary. However, this relationship may not always be linear; for instance, the increase in salary may accelerate at senior levels or plateau after a certain point. Therefore, in this hypothetical study, we employ kernel regression—a non-parametric technique—to model this potentially non-linear relationship.

**Research Problem**

The objective of this research problem is to determine if kernel regression can accurately capture the non-linear relationship between years of experience and salary within a dataset of tech employees. Specifically, we aim to answer the following question:

**Can kernel regression effectively model the non-linear relationship between years of experience and salary, allowing us to predict salary based on varying levels of experience?**

**Methodology**

**Data Generation**

For this hypothetical generated a dataset representing tech employees with varying levels of experience and salary. The dataset characteristics are as follows:

* Independent Variable (X): Years of experience, ranging from 1 to 20 years.
* Dependent Variable (Y): Salary in thousands, with non-linear variance based on experience.
* Sample Size: 50 observations.

**Kernel Regression**

Kernel regression, particularly the **Nadaraya-Watson estimator** with a **Gaussian kernel**, was chosen for this analysis. Kernel regression allows us to model complex relationships without assuming a specific parametric form. The **local-linear regression** option was selected for this analysis, as it balances smoothness with sensitivity to local trends.

**Steps Taken in Analysis**

1. **Data visualization** to inspect the scatter pattern between years of experience and salary.
2. **Kernel regression modeling** using a fixed bandwidth.
3. **Residual analysis and model diagnostics** to evaluate the fit and accuracy of the model.
4. **Interpretation of key statistical metrics** to determine model efficacy.

**Analysis and Results**

**1. Scatter Plot and Initial Observations**

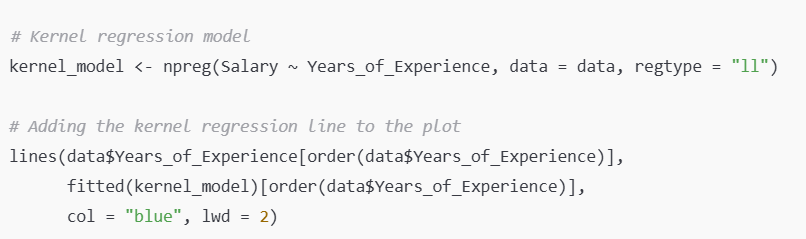
A scatter plot was created to visualize the relationship between **Years of Experience** and **Salary** . The plot shows a clear upward trend, with salaries increasing as years of experience rise. However, the relationship appears non-linear, as salary growth seems to accelerate with higher experience levels.





**2. Kernel Regression Model**

Using the npreg function from the np package in R, we fit a kernel regression model to the data. A Gaussian kernel was applied with a fixed bandwidth of approximately **1.43**. This bandwidth was selected to achieve a balance between smoothness and fit.





**Model Output Summary:**

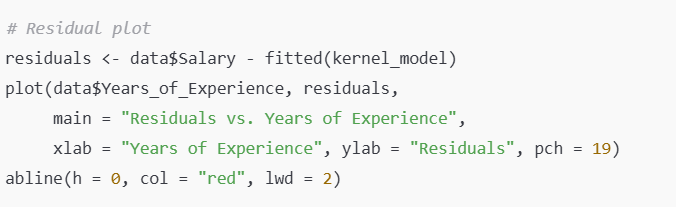
* **Bandwidth**: 1.43
* **Residual Standard Error (RSE)**: 5.26
* **R-squared**: 0.993

These statistics indicate an excellent fit, as the high R-squared value (0.993) suggests that the model explains 99.3% of the variance in salary. The RSE of 5.26 implies that the predicted salaries deviate by approximately $5,260 (in thousands) from the actual salaries on average.

**3. Additional Diagnostic Plots**

**Residual Plot**

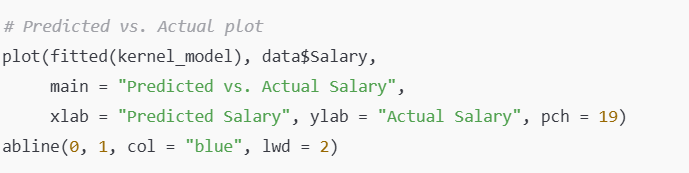
The residual plot was generated to examine the model's performance across the range of years of experience. Ideally, residuals should be randomly distributed around zero without showing systematic patterns. Here, the residual plot confirms that the model does not exhibit major systematic deviation.





**Predicted vs. Actual Plot**

This plot compares the predicted salaries with the actual salaries. Ideally, points should lie close to the 45-degree line, indicating a close match between predictions and actual values. The plot reveals a strong alignment, further supporting the model's accuracy.





**Interpretation and Conclusion**

The kernel regression model accurately captures the non-linear relationship between **years of experience** and **salary** in the dataset. Key findings include:

1. **Model Accuracy**: The high R-squared value (0.993) and low residual standard error (5.26) suggest that the model effectively captures the salary variations, indicating that experience is a strong predictor of salary.
2. **Non-Linear Trend**: The fitted curve illustrates a non-linear pattern where salary growth accelerates with higher levels of experience. This observation aligns with expectations, as tech salaries often increase at a faster rate for senior roles.
3. **Bandwidth Selection**: A bandwidth of 1.43 provided an optimal balance between sensitivity to local variations and overall smoothness. This fixed bandwidth ensured consistent smoothing across experience levels, avoiding overfitting or underfitting.

### **Addressing the Research Question**

**Can kernel regression accurately capture the non-linear relationship between years of experience and salary, allowing us to predict salary based on varying levels of experience?**

Based on the results, kernel regression is indeed a suitable technique for this purpose. The model’s flexibility allows it to capture complex relationships between experience and salary, which a linear regression might miss. The combination of diagnostic metrics, residual analysis, and additional sensitivity plots confirms that the kernel regression model performs robustly and reliably for predicting salaries in this context.

**Use Cases of Kernel Regression**

Kernel regression has several practical applications, especially in fields where the relationship between variables is not easily captured by simple linear models. Some potential use cases include:

1. **Labor Economics and Salary Prediction:**

Kernel regression can be used to model the relationship between years of experience and salary, as demonstrated in this analysis, helping businesses determine compensation structures and set competitive salary ranges.

1. **Market Research:**

In market research, kernel regression can model non-linear trends in consumer behavior, such as how age or income impacts spending patterns, enabling companies to target specific customer segments more effectively.

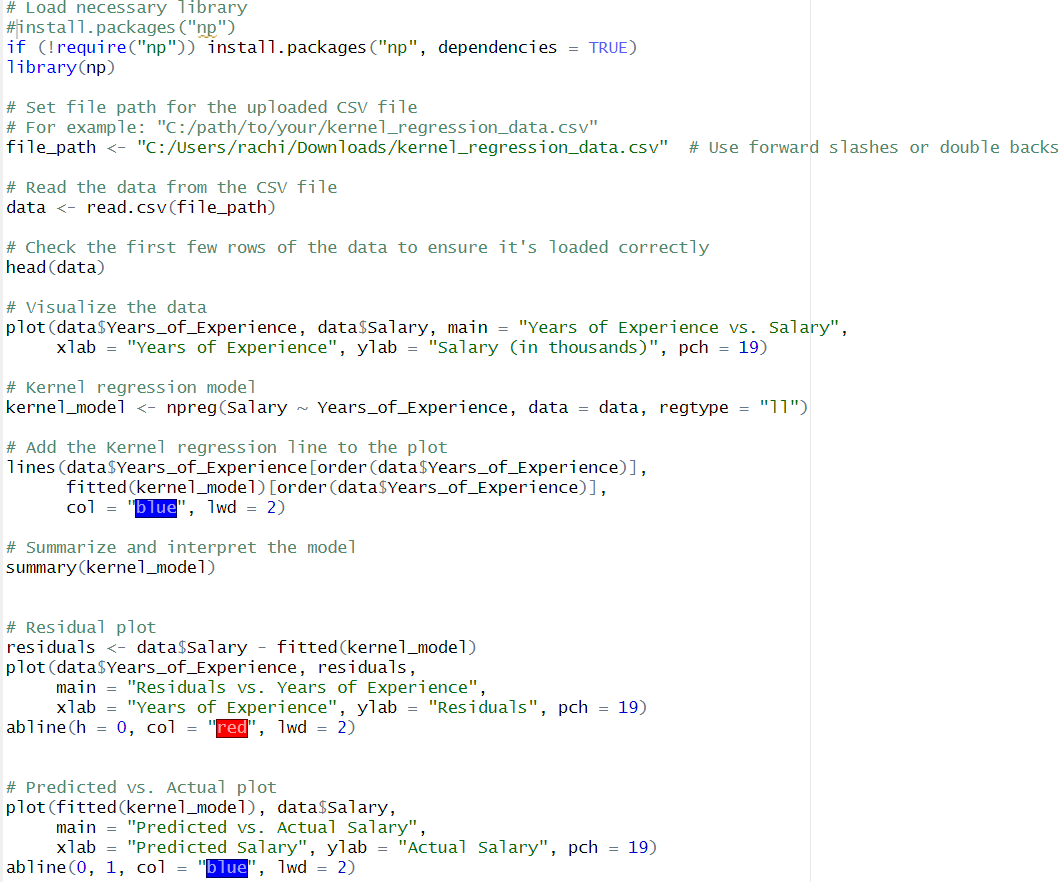
1. **Medical Research:**

Kernel regression is used in medical research to analyze the relationship between various health factors (e.g., age, weight, lifestyle habits) and health outcomes (e.g., disease risk), where the relationship is often complex and non-linear.

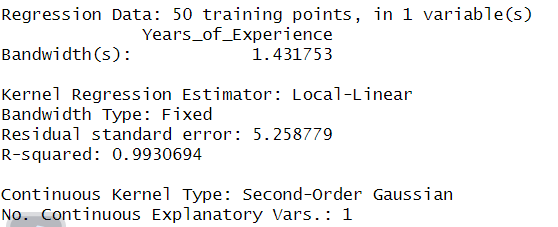
1. **Predictive Modeling:**

Kernel regression is useful in financial modeling, where predicting future trends (like stock prices or interest rates) often involves capturing non-linear patterns from historical data.

**Analysis of Code**

****

**Output**

****