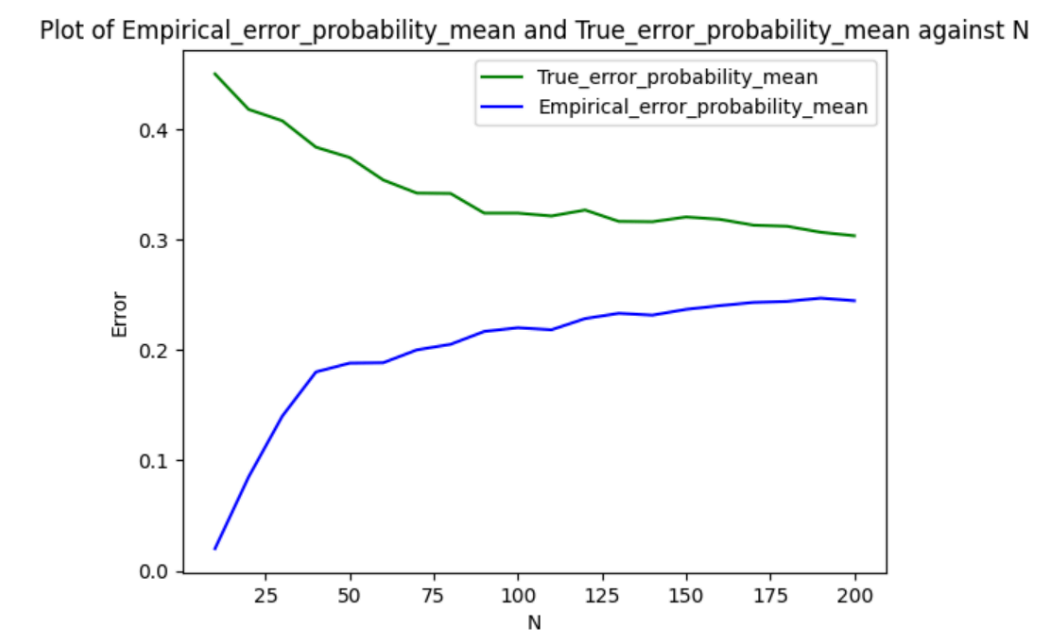
Lab1 : Report

2020csb1065
Aditya

Task: 1

In the case of empirical error, the loss function employed is the 0-1 loss function, while for the true error, it is the expected value of the 0-1 loss.

Subtask 1:

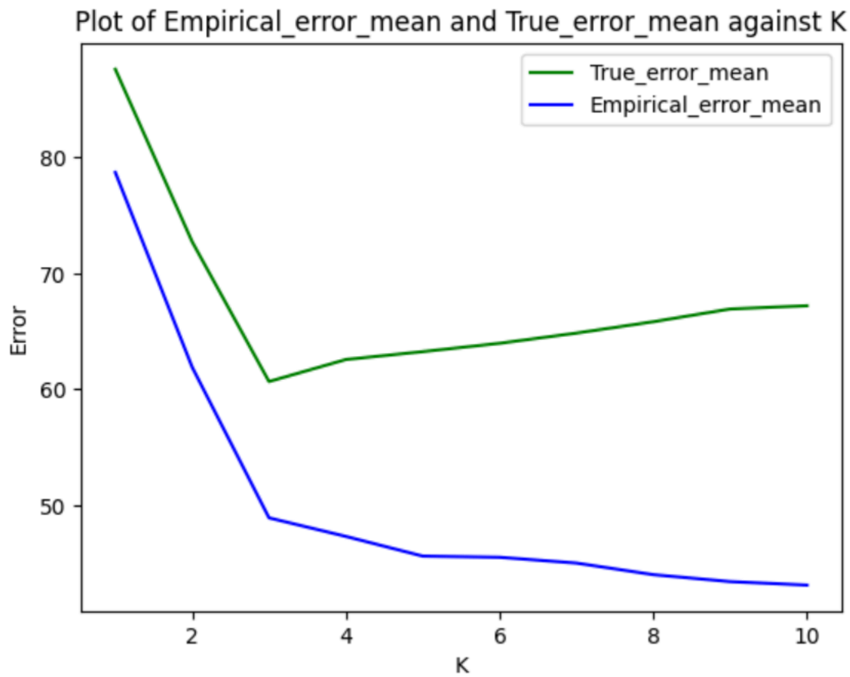


Results corresponding to k = 3

The Bayes optimal predictor, derived from the given probability distribution, assigns a value of 1 to x if it falls within the intervals [0, 0.2] ∪ [0.4, 0.6] ∪ [0.8, 1.0], and 0 otherwise. The true error of this predictor is calculated to be 0.26. Given the partition function provided, Raja will return the best possible tuple of partitions. Consequently, as the number of data points (N) increases, these partitions should converge towards the Bayes optimal predictor. It follows that the true error of the hypothesis should approach the true error of the Bayes optimal predictor. It's well-established that as the sample size increases, both the true error and empirical error tend to converge.

The plotted data aligns with this theory, showing that the true error of the hypothesis at k = 3 converges to 0.26, matching the true error of the Bayes optimal predictor. Additionally, as the sample size increases, both true and empirical errors exhibit convergence.

Subtask 2:



Results corresponding to N = 200

We observe the lowest empirical error at k = 10 and the lowest true error at k = 3. It's worth noting that the Bayes optimal hypothesis predicts 1 when x lies in the intervals [0, 0.2] ∪ [0.4, 0.6] ∪ [0.8, 1.0], and 0 otherwise. Additionally, according to the uniform convergence theorem, as we increase the sample size, the true error and empirical error tend to converge.

Despite achieving a low empirical error at k = 10, the optimal value of k remains 3, as it corresponds to the minimum true error. With a sufficiently large sample size, both empirical and true errors are expected to minimize at k = 3. Therefore, k* is 3, consistent with the Bayes optimal predictor.

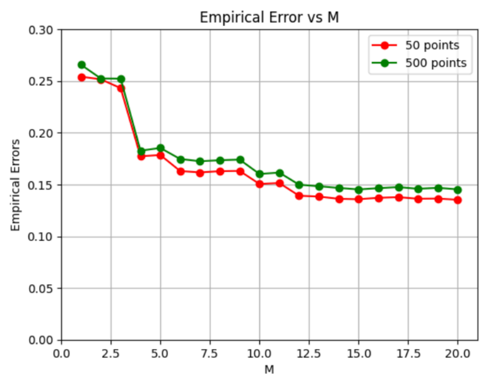
Task 2:

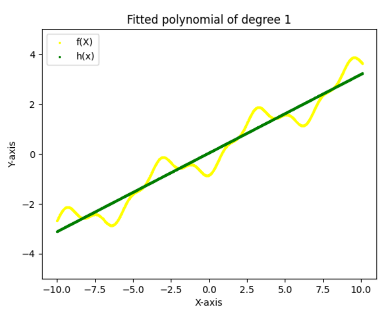
Gaussian Noise:

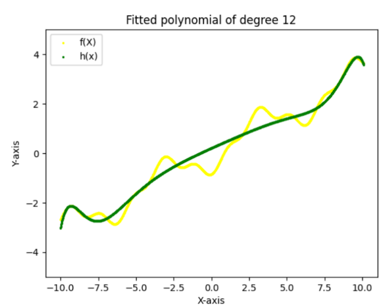
On increasing the degree

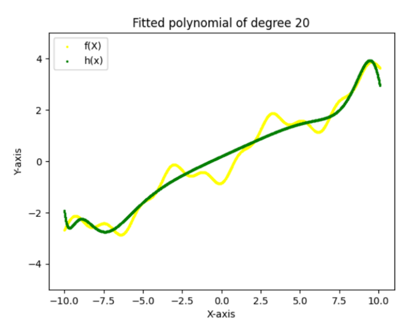
After reviewing the plots, it's apparent that the best-fit polynomial occurs around degree m=12. Despite achieving the lowest empirical error at degree 20, the polynomial is undoubtedly overfitted.

Conversely, the worst-fit polynomial is of degree 1, yielding the highest empirical error and failing to capture any fluctuations in the data.









Degree 1 **Degree 12** **Degree 20**

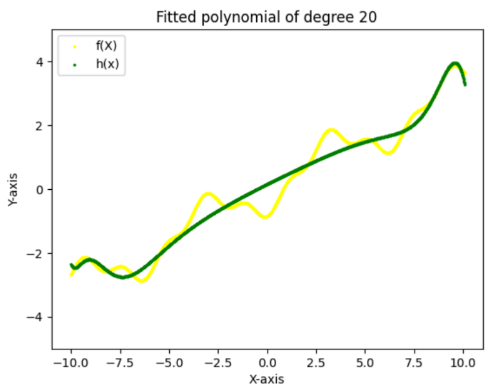
In the provided plots, the yellow line represents the true polynomial f(x), while the green line depicts the fitted polynomial h(x) when n = 50. Observing these plots, the degree-1 polynomial is visibly underfitted, performing poorly both on the sample and the true function, failing to capture any movements in f(x).

In contrast, the degree-12 polynomial shows improved performance by capturing variations and fluctuations in the function. Moreover, it exhibits a lower empirical error than degree-1. Thus, based on the plots, the degree-12 polynomial appears to be the best fit.

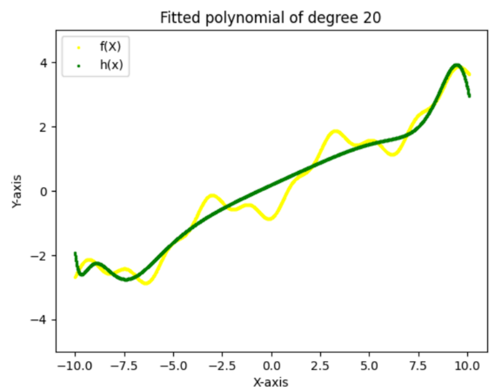
However, the degree-20 polynomial demonstrates overfitting tendencies. Although it boasts the lowest empirical error, a closer examination of the plot reveals that it captures noise in the data rather than the true fluctuations of the polynomial.

The above results are coherent with the theory, Increasing the degree of a polynomial in regression entails a bias-variance tradeoff. Lower degrees yield higher bias, oversimplifying relationships between features and the target. As degrees rise, flexibility increases, reducing bias by capturing complex patterns, yet raising variance due to sensitivity to data fluctuations, potentially leading to overfitting. Balancing these aspects is vital for models to generalize effectively. Methods such as cross-validation and regularization aid in determining the optimal degree, ensuring models capture underlying patterns without overfitting.

On increasing the number of datapoints



N = 50



N = 500

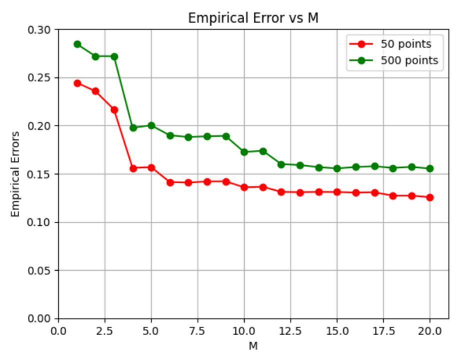
The displayed plots showcase fitted polynomials of degree 20, with sample sizes of 50 and 500 data points. Overfitting occurs more prominently with smaller sample sizes, while with larger sample sizes, there is a decrease in overfitting. The risk of overfitting diminishes as the number of data points increases, especially noticeable at the extremes of the x-axis.

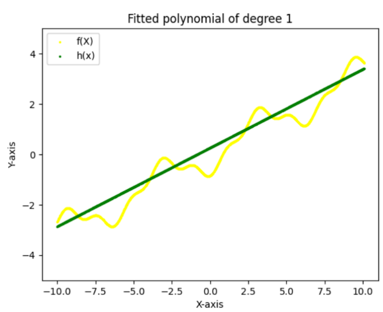
Poisson Noise

On Increasing the degree

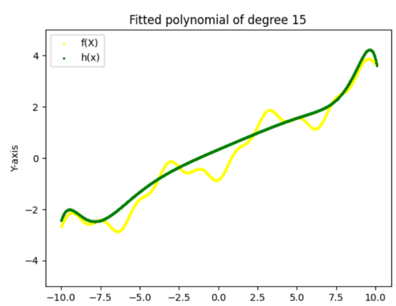
After reviewing the plots, it's apparent that the best-fit polynomial occurs around degree m=15. Despite achieving the lowest empirical error at degree 20, the polynomial is undoubtedly overfitted.

Conversely, the worst-fit polynomial is of degree 1, yielding the highest empirical error and failing to capture any fluctuations in the data.

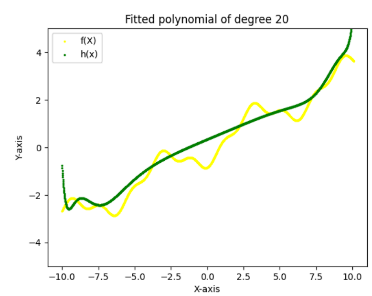




Degree 1



Degree 15



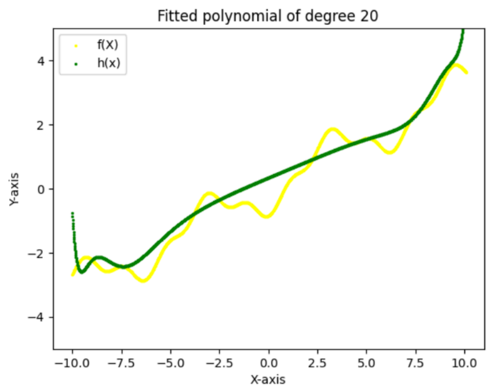
Degree 20

In the provided plots, the yellow line represents the true polynomial f(x), while the green line depicts the fitted polynomial h(x) when n = 50. Observing these plots, the degree-1 polynomial is visibly underfitted, performing poorly both on the sample and the true function, failing to capture any movements in f(x).

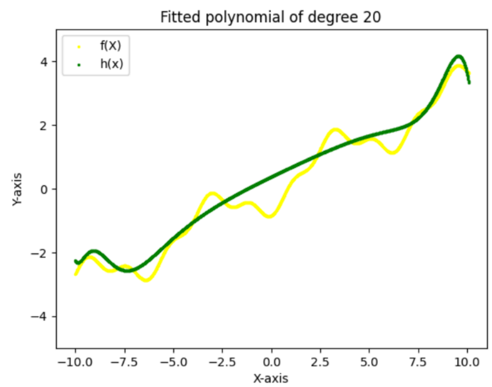
In contrast, the degree-15 polynomial shows improved performance by capturing variations and fluctuations in the function. Moreover, it exhibits a lower empirical error than degree-1. Thus, based on the plots, the degree-15 polynomial appears to be the best fit.

However, the degree-20 polynomial demonstrates overfitting tendencies. Although it boasts the lowest empirical error, a closer examination of the plot reveals that it captures noise in the data rather than the true fluctuations of the polynomial.

On increasing the number of datapoints



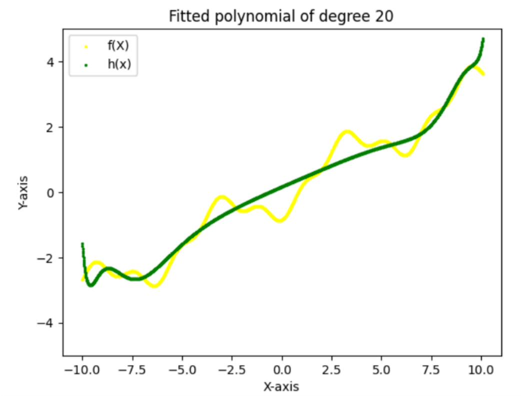
N = 50

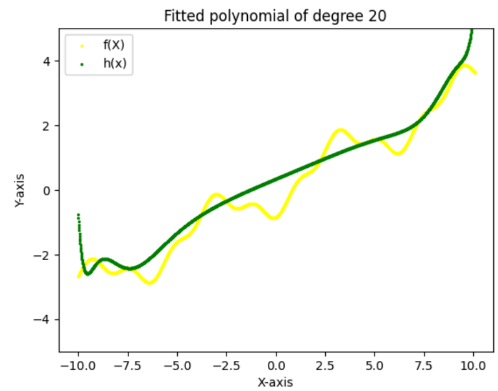


N = 500

The displayed plots showcase fitted polynomials of degree 20, with sample sizes of 50 and 500 data points. Overfitting occurs more prominently with smaller sample sizes, while with larger sample sizes, there is a decrease in overfitting. The risk of overfitting diminishes as the number of data points increases, especially noticeable at the extremes of the x-axis.

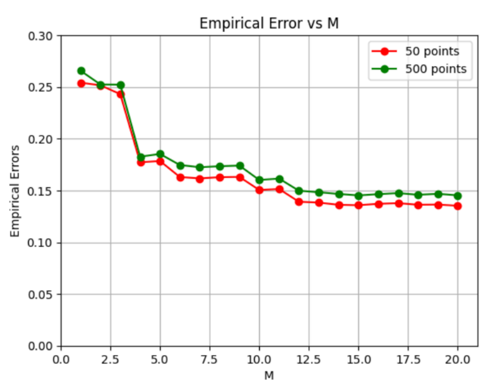
Comparing noise

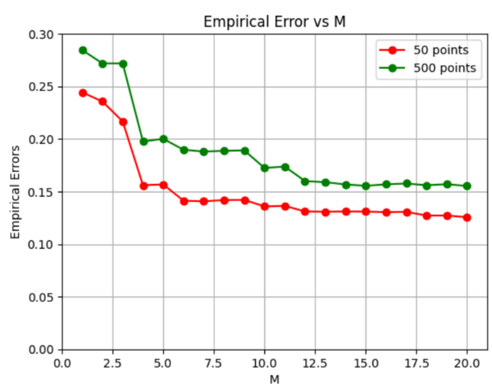




Gaussian noise **Poisson noise**

The graph above illustrates the scenario when the sample size N =50 and the degree of the polynomial is 20. With Poisson noise, there's a notable tendency for overfitting, particularly evident at the extremes. The fitted polynomial encounters challenges in accurately capturing the underlying trends in the data due to significant fluctuations introduced by the Poisson noise.



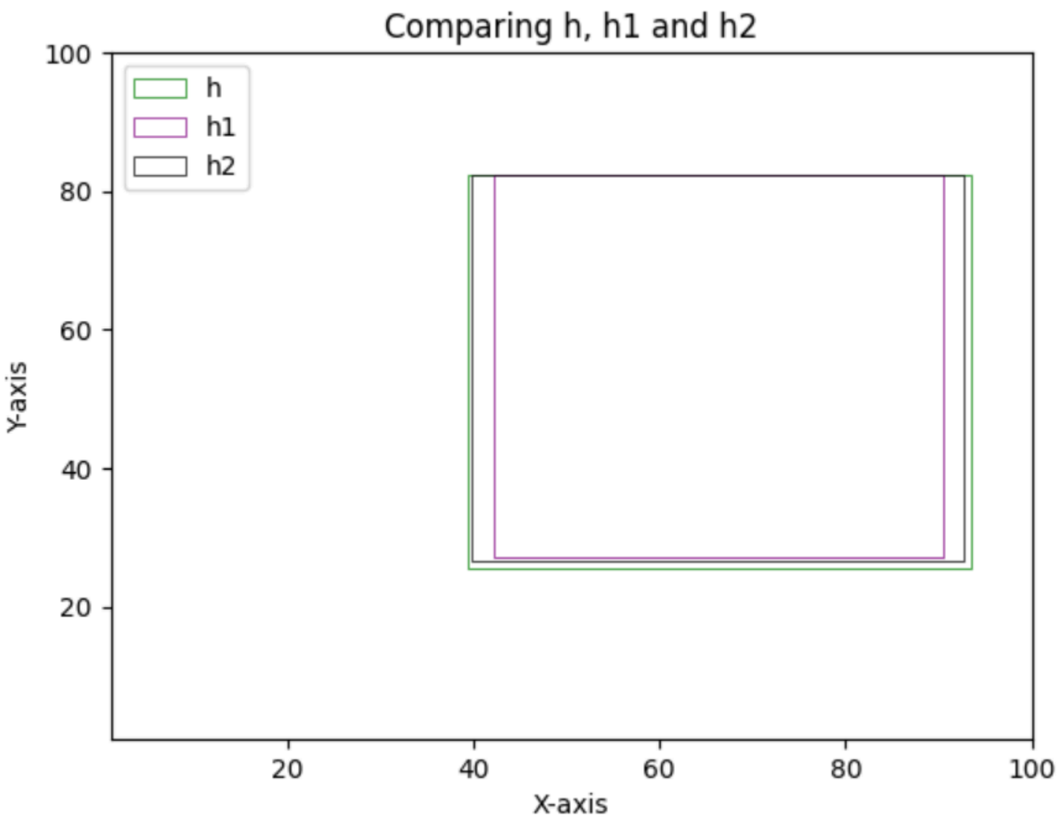


Gaussian noise **Poisson noise**

With Poisson noise, the empirical error increases notably as the sample size is augmented from N=50 to N=500 compared to Gaussian noise.This implies that Poisson noise introduced erratic fluctuations that are more challenging for the model to handle. As the sample size grows, the impact of these fluctuations becomes more pronounced, resulting in a notable rise in empirical error.

The empirical error is higher with Poisson noise compared to Gaussian noise when the sample size remains constant. This discrepancy arises due to the unique characteristics of Poisson noise, which is characterized by discrete and non-negative values. These properties can introduce greater variability and fluctuations in the data, making it more challenging for the model to accurately capture the underlying patterns.

Task: 3



Plot of h, h1 and h2

We understand theoretically that axis-aligned rectangles are PAC-learnable, which aligns with the experimental findings discussed above. When generating hypotheses based on 100 sample points, we obtained hypothesis h1, and when using 200 sample points, we derived hypothesis h2. This suggests that with larger samples, the generated hypotheses tend to converge towards the true hypothesis. This convergence is clearly observable as h2 begins to converge towards h.