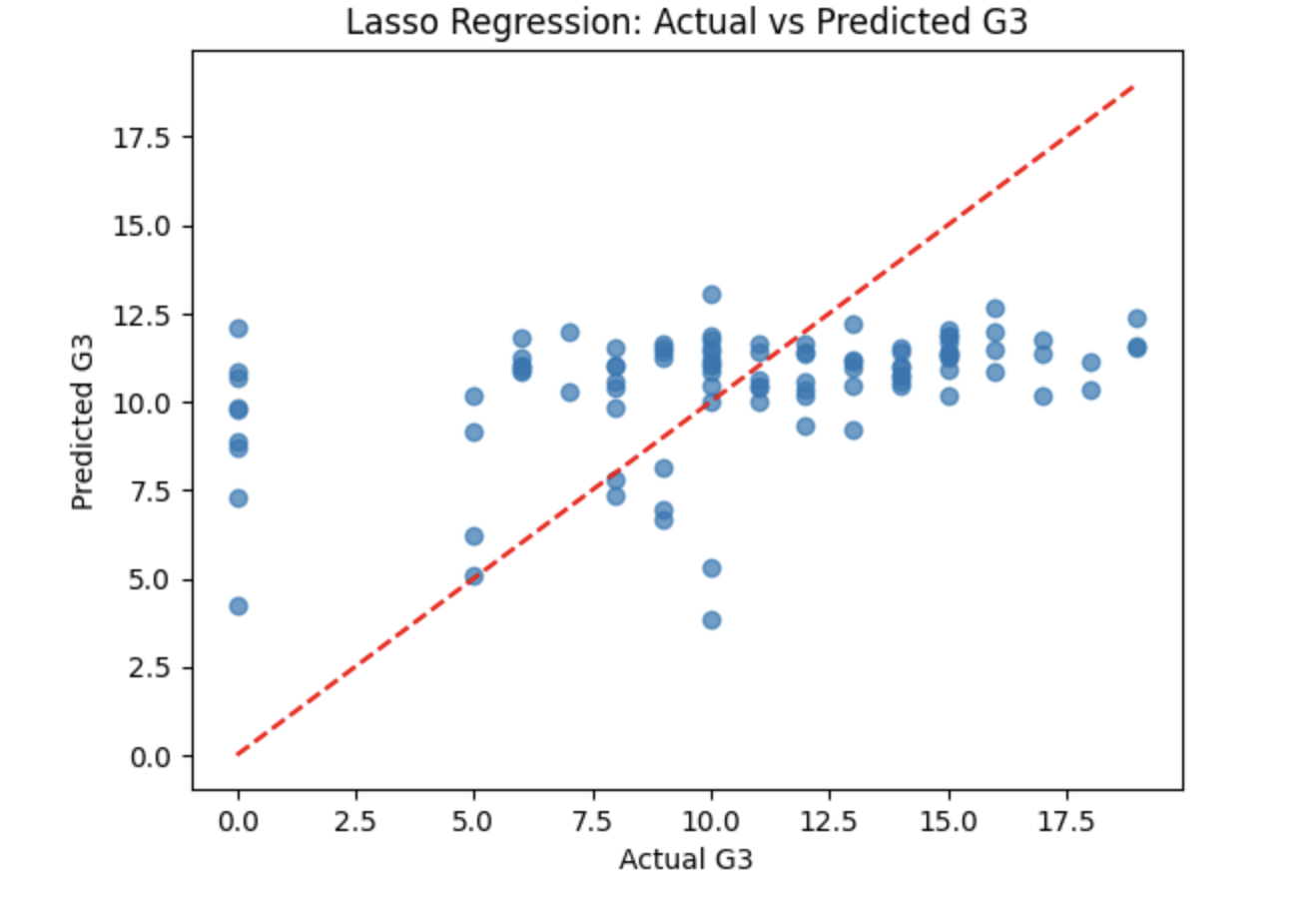
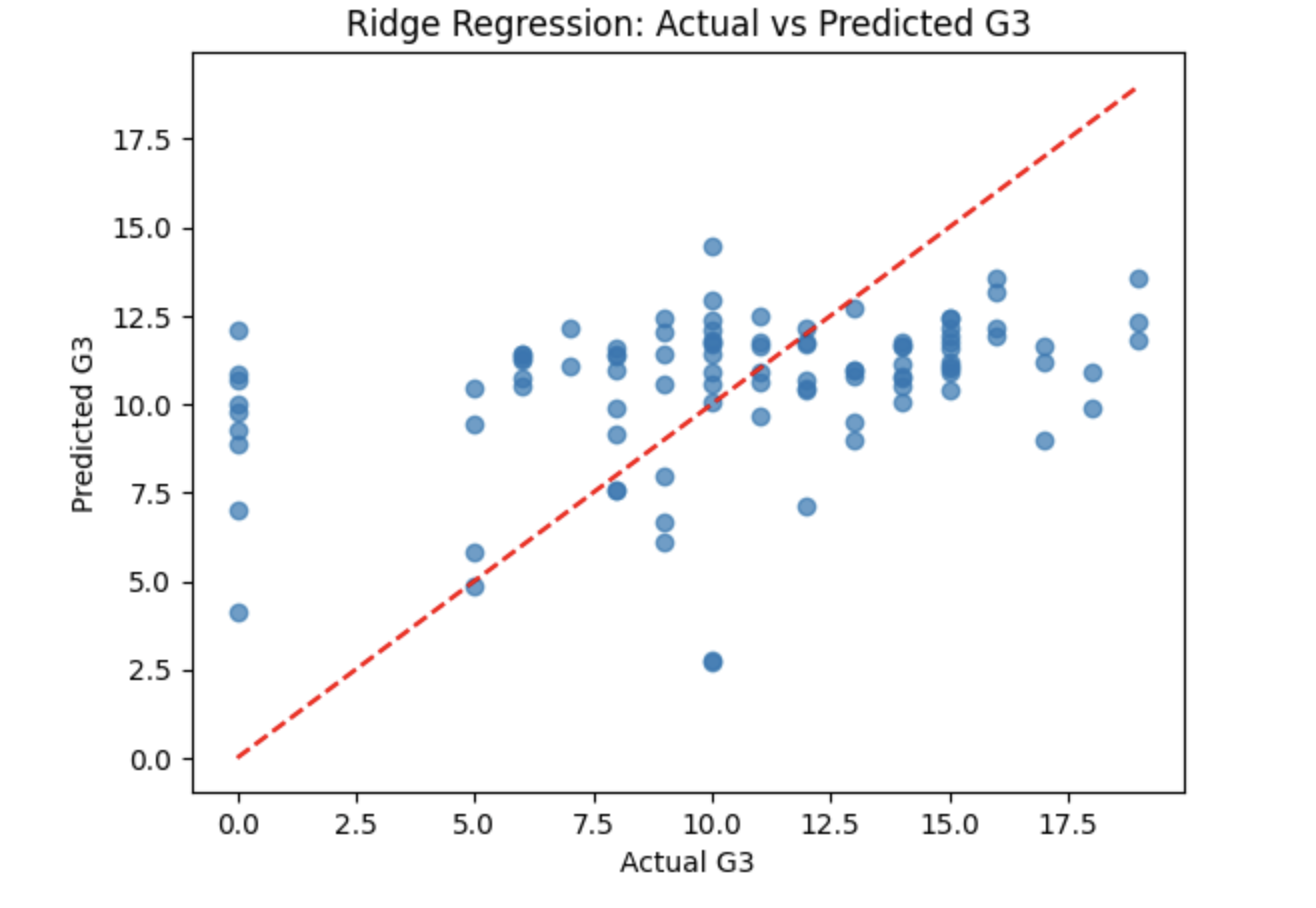
### **Ridge and Lasso Regression**

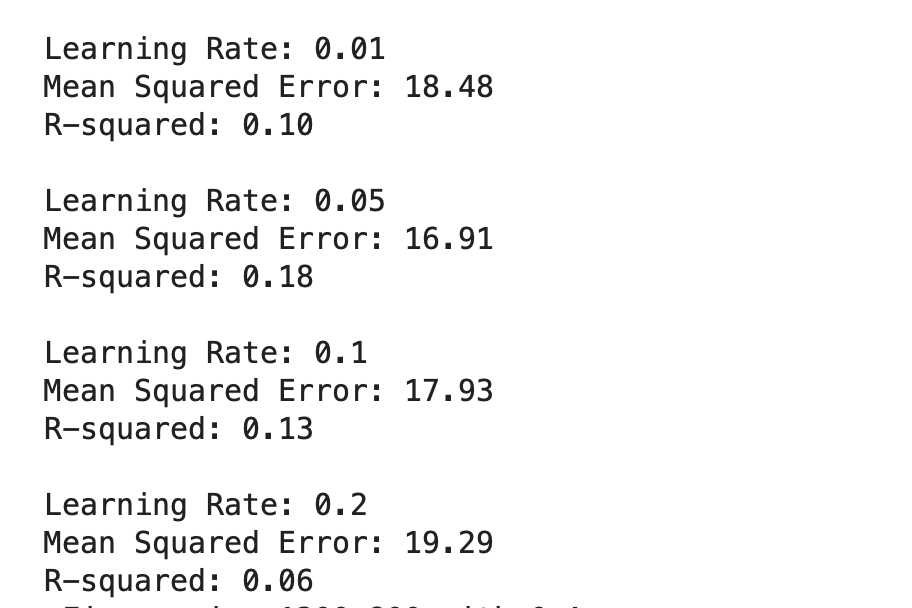
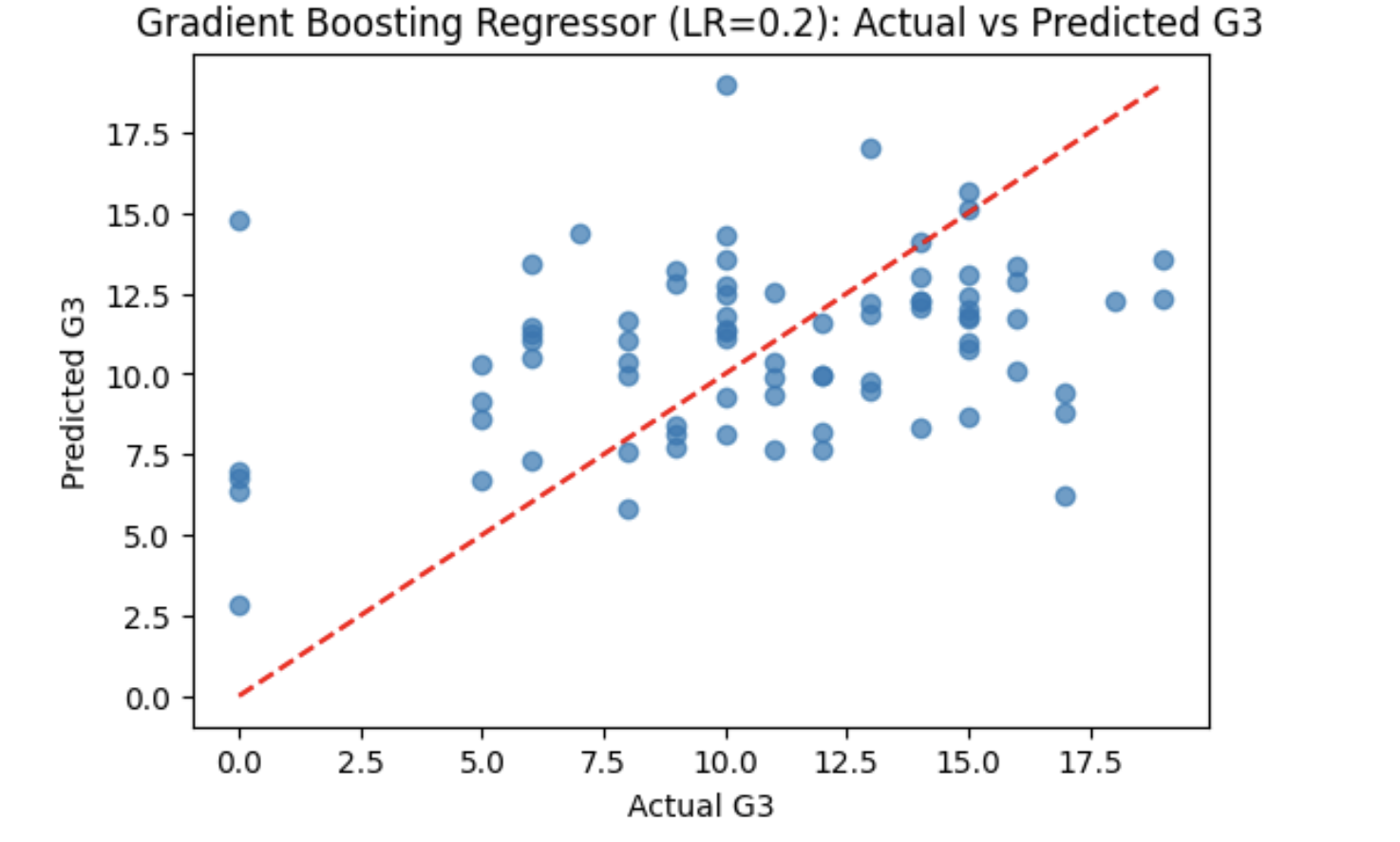
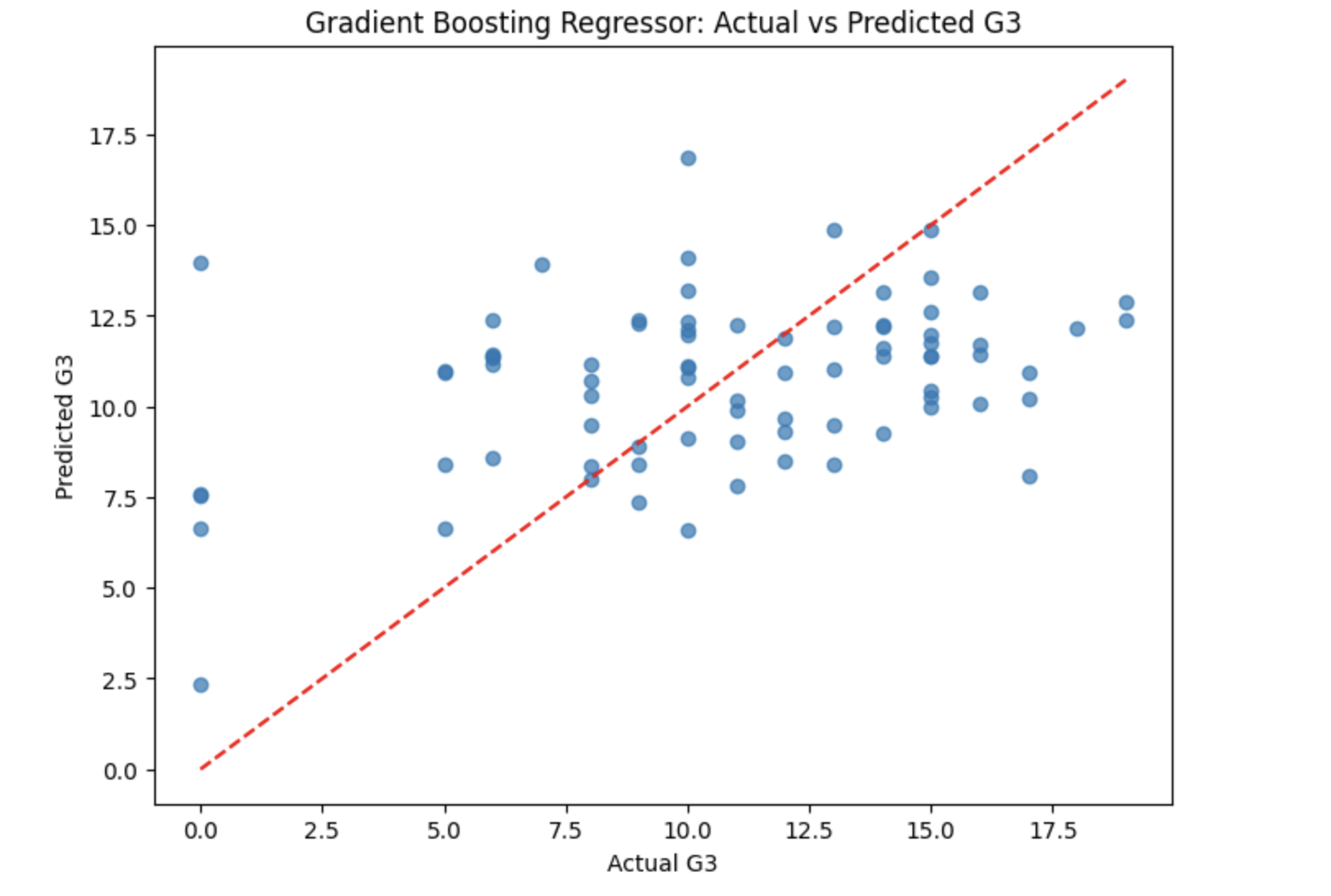
In this study, Ridge and Lasso Regression were used to predict students' final grades based on a variety of features, including parental education, study habits, extracurricular involvement, and support from family and school. These methods were chosen because of their ability to handle multicollinearity and potentially select the most important predictors.

* **Ridge Regression**: Ridge Regression was applied to address potential multicollinearity issues within the features. It adds an L2 penalty term to the loss function, which helps shrink coefficients, reducing model complexity and variance. We used various regularization strengths (alphas), ranging from 0.01 to 100. The performance of the Ridge Regression model was moderate, with R-squared values ranging between 0.12 and 0.15 across different alpha values. This indicates that the model was able to explain only a limited portion of the variance in the students' final grades. Despite its ability to handle correlated predictors well, the overall performance improvement over the baseline linear regression was marginal, suggesting that there might be missing features or non-linear relationships that the model could not capture effectively.
* **Lasso Regression**: Lasso Regression was also employed as an alternative to Ridge Regression due to its L1 penalty term, which promotes sparsity in the coefficients. This characteristic allows Lasso to effectively perform feature selection by driving some coefficients to zero, thereby excluding less relevant features. We experimented with various alpha values and found that the best performance was obtained with an alpha of 0.1, which yielded an R-squared value of 0.16. While the improvement over Ridge Regression was not substantial, it suggested that some features in our dataset might not be contributing meaningfully to the model, and Lasso was able to identify those. The L1 regularization led to a slightly more interpretable model, which might be beneficial for understanding which factors were the most influential, but the predictive power remained relatively low. This suggests that, although Lasso is capable of feature selection, the current dataset's features might still be insufficient for accurate grade prediction.

### **Gradient Boosting Regression**

In addition to Ridge and Lasso, we utilized Gradient Boosting Regression, motivated by findings from previous studies such as Martins et al. (2021) and Goyal & Kumar (2023), which demonstrated the effectiveness of Gradient Boosting in educational settings. Gradient Boosting is an ensemble method that builds a series of decision trees sequentially, with each new tree trying to correct the errors made by the previous ones. This approach is particularly powerful for capturing complex relationships in the data.

* **Gradient Boosting Regressor**: We experimented with different learning rates (0.01, 0.05, 0.1, and 0.2) to assess the impact on model performance. The model with the lowest Mean Squared Error (MSE) of 17.93 and an R-squared value of 0.13 was achieved with a learning rate of 0.05. Although Gradient Boosting showed a slight improvement over Ridge and Lasso Regression, the overall fit was still weak. The R-squared values indicated that only a small fraction of the variability in students' final grades was captured by the model. Moreover, residual plots revealed inconsistent patterns, suggesting that the model failed to capture key factors influencing student outcomes.  
  The use of Gradient Boosting was inspired by its ability to model complex non-linear relationships and handle feature interactions effectively. However, our results suggest that while Gradient Boosting can capture some of these complexities better than linear models, it still struggled with the features available in our dataset. This outcome is in line with the findings of Martins et al. (2021), who also used Gradient Boosting to predict student dropout and found that, despite its advantages, the model's performance was limited by the quality and completeness of the features. Similarly, Goyal & Kumar (2023) faced challenges with data imbalance, which also likely played a role in our study's limited model performance. These limitations highlight the difficulty of predicting educational outcomes using a limited set of features and the need for richer data to improve predictive accuracy.



In summary, Ridge and Lasso Regression provided some level of feature insight, with Lasso demonstrating slight feature selection capabilities, while Gradient Boosting showed a marginal improvement in predictive performance. Nevertheless, the relatively low R-squared values across all models indicate that predicting student performance accurately remains a challenging task. This suggests a need for either more sophisticated modeling techniques or the inclusion of additional, more informative features to fully capture the complex factors that influence student success.