**MACHINE LEARNING PROBLEM SET 2**

**EMPIRICAL PROBLEMS**

**2. ANALYSIS**

**(f)**

Summary Table of Model Estimates:

| Model | Lambda | MSE |
| --- | --- | --- |
| Ridge | 0 | 20235.48 |
| Ridge | 5 | 20220.05 |
| Ridge | 10 | 20226.91 |
| Lasso | 0 | 20235.48 |
| Lasso | 5 | 20408.40 |
| Lasso | 10 | 20657.49 |

As the penalty parameter (lambda) increases, Ridge regression initially improves with a slight decrease in MSE from 20235.48 to 20220.05 when lambda rises from 0 to 5, but over-regularization at lambda = 10 causes the MSE to increase to 20226.91. This suggests that moderate penalties can enhance Ridge performance, while excessive penalization may degrade it. In contrast, Lasso regression shows a more pronounced sensitivity to the penalty, with MSE increasing from 20235.48 at lambda = 0 to 20408.40 at lambda = 5, and further to 20657.49 at lambda = 10, indicating deteriorating predictive performance as penalties increase.

Ridge regression tends to be more stable as the penalty increases, as it shrinks coefficients without eliminating any, which helps with multicollinearity. A moderate penalty (like lambda = 5) may reduce overfitting while maintaining predictive accuracy. In contrast, Lasso regression, which shrinks some coefficients to zero for feature selection, can lead to a loss of important information as lambda increases, resulting in poorer predictive performance as seen in the rising MSE values.

In summary, both models demonstrate the importance of choosing an appropriate penalty parameter. Ridge appears more robust against higher penalties than Lasso in this analysis.

**(g)**

The MSE results indicate that the linear model performed exceptionally well with a low MSE of 11.28, effectively capturing key relationships but possibly oversimplifying the price dynamics. While specific MSE values for the Polynomial and Interaction Model and the BIC-Selected Model were not provided, they likely had higher MSEs due to increased complexity. Ridge regression achieved a best MSE of 20220.03, benefiting from regularization to manage multicollinearity, while Lasso regression had a slightly worse MSE of 20232.71, indicating its feature selection may not always enhance performance. Overall, while the linear model is interpretable and accurate, Ridge regression may offer better generalization for more complex datasets, highlighting the trade-off between model complexity and predictive performance.

**(h)**

Given the task of predicting Airbnb listing prices, Ridge regression is the most appropriate linear supervised learning procedure, especially considering the sample size and covariate structure. Ridge effectively handles multicollinearity among predictors, which is common in Airbnb data with correlated features like the number of accommodates and beds. Its ability to regularize estimates helps prevent overfitting, making it suitable for larger sample sizes. Additionally, while primarily focused on linear relationships, Ridge can accommodate polynomial terms or interactions, allowing for flexibility without sacrificing predictive performance. Overall, Ridge regression balances accuracy and interpretability, making it an ideal choice for this prediction task.

**3. (c)**

The changes in MSE across different models can be attributed to several factors. Simpler models, like linear regression, tend to have lower MSE due to their straightforward interpretation and fewer parameters, making them less prone to overfitting. As complexity increases with polynomial and interaction terms, MSE may rise if the model captures noise rather than the underlying data structure. Regularization in Ridge and Lasso regression helps reduce MSE by controlling multicollinearity and overfitting, leading to better generalization. Cross-validation further optimizes MSE by tuning model parameters to balance fit and robustness. Ultimately, MSE variations reflect the interplay between model complexity, regularization effects, and data variability.

**CONCEPTUAL PROBLEMS**

1. If is a linear and unbiased estimator of it effectively captures the relationship between and with minimal systematic error, resulting in a lower mean squared error (MSE) compared to ().

* estimates a constant mean, ignoring the true functional form , which leads to higher bias and MSE.
* s a quadratic model that may introduce additional variance if overfitted or misspecified, resulting in a higher MSE than the linear model.

Thus, , being linear and unbiased, is expected to yield the lowest MSE among the three models, accurately reflecting the true data-generating process.

1. To derive the relationships between the coefficients of the three models at :

**Model 1**: ​. This model estimates the mean of as suggesting .

**Model 2**: . Since it's unbiased, we expect, leading to.

**Model 3**:. To remain unbiased, ​ is likely close to 0, , and should ideally be 0 as the quadratic term is unnecessary for the true function.

In summary, at , the coefficients should satisfy and ideally being 0 to accurately approximate the linear function.