Report on Regression on Bike Sharing Demand

Problem Statement. The objective is to predict hourly bike rental demand using regression models on the Bike Sharing Demand dataset and to select the best model strictly based on test-set Mean Squared Error (MSE) and coefficient of determination (R^2).

Objective. We analyze how increasing model nonlinearity affects predictive performance by comparing a linear baseline, polynomial regression models of degree $d = 2, 3, 4$ *without interaction terms*, and a quadratic polynomial model of degree $d = 2$ *with interaction terms*, to determine the optimal bias–variance tradeoff.

Preprocessing and Target Stabilization. Time-dependent cyclic features sin(hour) and cos(hour), seasonal indicators, and meteorological variables are used. Numerical features are standardized and categorical features are one-hot encoded without leakage. The target is log-transformed:

$$y_{\text{log}} = \log(1 + y),$$

and debiased using the smearing estimator:

$$\hat{y}_{\text{final}} = \mathbb{E}\left[\frac{y}{\hat{y}}\right]_{\text{train}} \cdot \hat{y}_{\text{test}}.$$

Model Definitions and Interpretation

1. Linear Regression (Baseline).

$$\hat{y} = w^\top x$$

This model assumes a purely linear relationship between features and demand. It has high bias and is unable to model nonlinear seasonality and temperature effects.

2. Polynomial Regression ($d = 2$, No Interactions).

$$\hat{y} = w_0 + \sum_j w_j x_j + \sum_j w_j^{(2)} x_j^2$$

This captures basic curvature in demand with respect to temperature and environmental variables, while maintaining low variance.

3. Polynomial Regression ($d = 3$, No Interactions).

$$\hat{y} = w_0 + \sum_j w_j x_j + \sum_j w_j^{(2)} x_j^2 + \sum_j w_j^{(3)} x_j^3$$

This model captures higher-order nonlinear saturation effects and achieves the best generalization by balancing bias and variance.

4. Polynomial Regression ($d = 4$, No Interactions).

$$\hat{y} = w_0 + \sum_{k=1}^{4} \sum_j w_j^{(k)} x_j^k$$

This introduces excessive curvature and increases variance despite strong regularization, leading to mild overfitting.

5. Quadratic Polynomial with Interaction Terms ($d = 2$).

$$\hat{y} = w_0 + \sum_j w_j x_j + \sum_{j \le k} w_{jk} x_j x_k$$

This model explicitly captures both individual feature effects and pairwise feature interactions (e.g., temperature–humidity coupling). While it increases representational power, it also raises model variance due to the expansion in feature space.

All polynomial models are trained using Ridge regularization:

$$\min_w \frac{1}{n} \sum_{i=1}^{n} \Big(y_i - \big(w_0 + w^\top x_i \big) \Big)^2 + \lambda \|w\|^2$$

where λ is the regularization parameter selected using cross-validated RidgeCV.

Results (Test Set).

Model	**MSE**	R^2
Linear Regression	39458.66	0.1666
Polynomial $d = 2$ (No Interactions)	32461.25	0.3144
Polynomial $d = 3$ (No Interactions)	**27292.57**	**0.4236**
Polynomial $d = 4$ (No Interactions)	28580.39	0.3964
Polynomial $d = 2$ (With Interactions)	27522.78	0.4187

The best performing model is the **degree $d = 3$ polynomial without interaction terms** with Ridge regularization parameter $\alpha = 66$.

Final Conclusion. The linear model underfits due to excessive bias. The degree-2 polynomial improves performance by introducing curvature. The degree-3 polynomial achieves the optimal bias–variance tradeoff and provides the best test performance. The degree-4 and quadratic interaction models increase variance and do not surpass the cubic model. Hence, the optimal model for bike demand forecasting in this study is the cubic non-interaction polynomial.

Final Selected Model: Polynomial Regression of Degree $d = 3$ (No Interactions, Ridge-Regularized)

Done by:
Akshaya – BT2024215
Geethika – BT2024139