COMP551-A2-Report

Jawdat Al-Jabi, Tal Smith, Steven Thao

October 2025

Abstract

This assignment examines model complexity, regularization, and evaluation using synthetic data and from scratch implementations. We first fit linear regression with Gaussian basis functions of increasing complexity $D \in 0, 5, \ldots, 45$ to a noisy nonlinear target, plotting bases, fitted curves, and train/validation sum of squared errors (SSE) to select D and illustrate the transition from underfitting to overfitting. We then study the bias-variance trade-off by repeating the full pipeline 10 times with new noise realizations, overlaying all fits, the ground truth, and the mean model, and averaging train/test errors to separate systematic bias from variance. Next, we incorporate ℓ_1 (Lasso) and ℓ_2 (Ridge) regularization into the $D = 45$ model, perform 10-fold cross-validation over λ, and plot the train and validation error alongside a bias-variance decomposition versus λ, highlighting how λ trades variance for bias and how ℓ_1 encourages sparsity while ℓ_2 produces smooth shrinkage. Finally, we visualize loss contours and gradient descent trajectories for ℓ_1 and ℓ_2 on linear synthetic data to connect optimization geometry with regularization effects. Throughout, we emphasize reproducible data generation, consistent plot ranges for fair comparisons, and clear reporting aligned with the required deliverables. Together, the experiments provide a cohesive view of complexity control and regularization in linear models and practical guidance for selecting D and λ in noisy settings.

1 Introduction

This assignment investigates how model complexity, regularization, and evaluation shape generalization in linear models are used using controlled synthetic data and from scratch implementations in Python. We study linear regression with nonlinear Gaussian basis functions, repeat experiments to analyze the bias-variance trade-off, introduce ℓ_1 (Lasso) and ℓ_2 (Ridge) regularization with 10-fold cross-validation, and visualize loss landscapes with gradient–descent trajectories to link optimization geometry to regularization effects. Concretely, we address four tasks:

- **Task 1: Linear Regression with Gaussian Bases.** We generate $N = 100$ samples from $y(x) = (\log x + 1)\cos x + \sin(2x) + \epsilon$ on $[0, 10]$ with $\epsilon \sim \mathcal{N}(0, 1)$, construct $D \in 0, 5, \ldots, 45$ Gaussian basis functions with μ_i linearly spaced across the input range and $\sigma = 1$, and fit linear models for each D. We plot the basis functions, fitted curves alongside the noisy data and ground truth, and compute the sum of squared errors (SSE) on the train and validation splits to select D that minimizes the validation SSE. We discuss the transition from underfitting to overfitting and the value of validation for model selection.

- **Task 2: Bias–Variance with Multiple Fits.** We repeat the full Task 1 pipeline ten times with fresh noise realizations. For chosen values of D, we overlay the ten fitted functions (green), the ground truth (blue), and the mean fit (red), and we plot average train and test errors across repetitions. These visualizations and aggregates support a qualitative and quantitative discussion of bias and variance.

- **Task 3: Regularization with Cross-Validation.** We augment the $D = 45$ model with ℓ_1 and ℓ_2 penalties, sweep λ on a logarithmic grid, and perform 10-fold cross-validation. We report train and validation errors versus λ and compute bias–variance decompositions as functions of λ to show how regularization trades variance for bias and how ℓ_1 encourages sparsity while ℓ_2 yields smooth shrinkage. We select λ^* that minimizes validation error and justify the choice.

- **Task 4: Effect of ℓ_1 vs. ℓ_2 on Loss.** Using linear synthetic data $y = -3x + 8 + 2\epsilon$ with $\epsilon \sim \mathcal{N}(0, 1)$ and $N = 30$, we plot loss contours for ℓ_1 and ℓ_2 objectives at multiple λ values and overlay gradient–descent trajectories. These figures illustrate optimization geometry, shrinkage, and sparsity effects under different regularization strengths.

2 Task 1

2.1 Data Generation:

We generated $N = 100$ data points uniformly from $[0, 10]$ and evaluated

$$y(x) = (\log x + 1)\cos x + \sin(2x) + \epsilon,$$

with $\epsilon \sim \mathcal{N}(0, 1)$ to add noise. This figure shows how the sampled points deviate from the smooth underlying function, illustrating why model complexity matters for fitting.

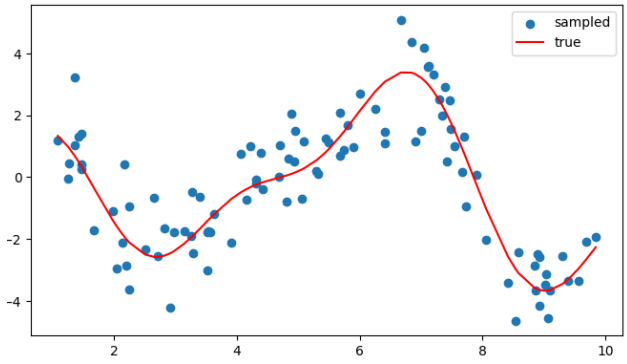

Figure 1: Clean vs Noisy Data

2.2 Gaussian Basis Functions and Model Fitting:

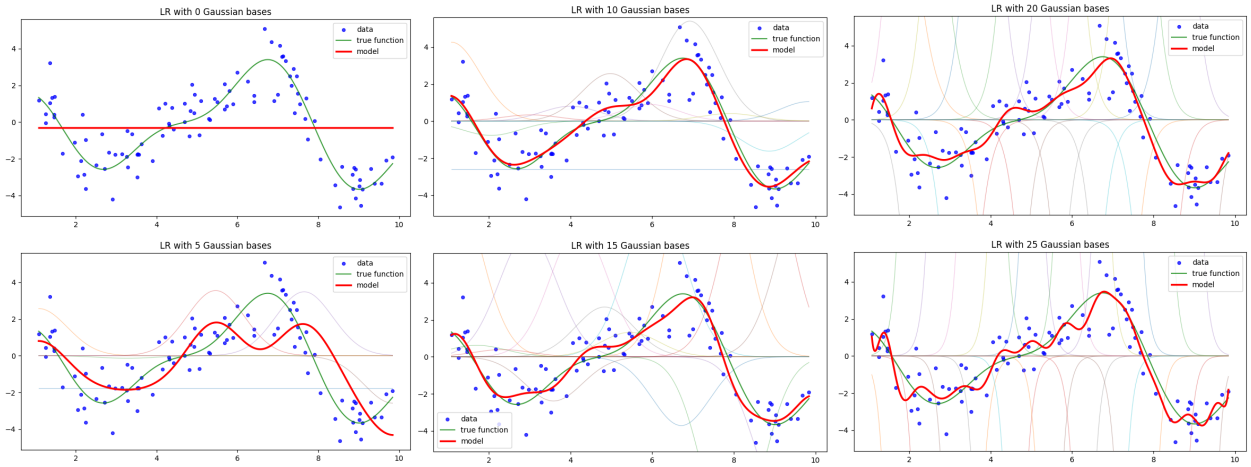

Figure 2: Fitted Model Plots for D = 0, 5, 10, 15, 20, 25

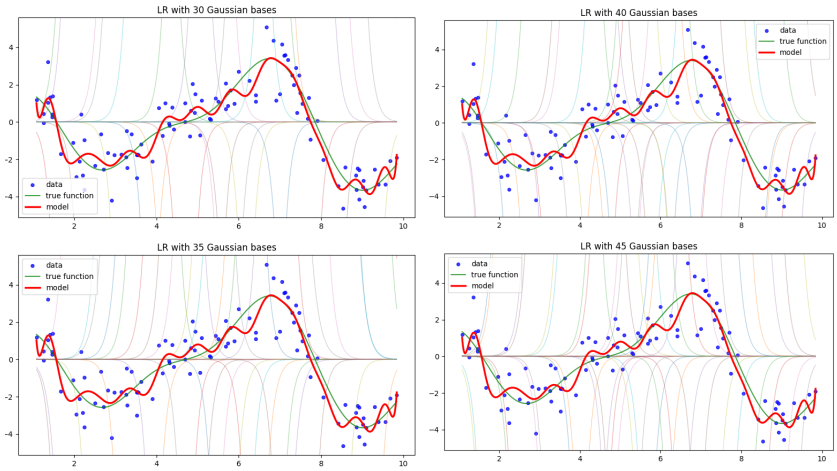

Figure 3: Fitted Model Plots for D = 30, 35, 40, 45

We applied Gaussian basis functions of the form

$$\phi(x; \mu, \sigma) = \exp\Big(-\frac{(x-\mu)^2}{\sigma^2}\Big),$$

with $\sigma = 1$ and μ_i equally spaced across $[0, 10]$. As D increases, the model can represent increasingly complex patterns. At low D, the model underfits and fails to capture oscillations. At intermediate D, the model aligns closely with the signal. At high D, the model begins to overfit noise.

2.3 Model Selection:

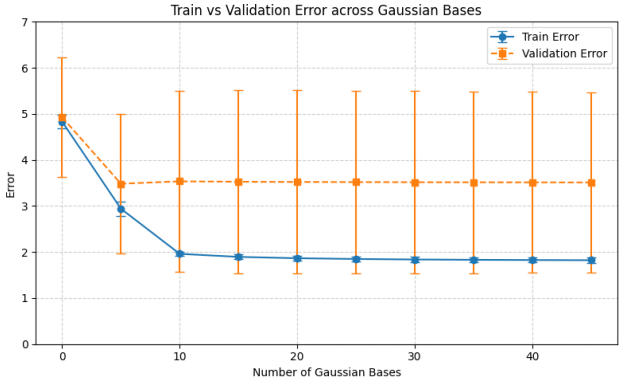

Figure 4: Train and validation SSE as a function of the number of Gaussian bases D.

We computed $\text{SSE}_{\text{train}}$ and SSE_{val} for each model. Training error decreases monotonically as D increases. Validation error reaches its minimum around $D = 10$, while the validation error at $D = 5$ is nearly identical within one standard deviation. Because the model with $D = 5$ achieves similar performance with lower complexity, we selected $D^* = 5$ as the optimal number of Gaussian bases. This choice balances model simplicity and validation performance.

3 Task 2

3.1 Multiple Fits Visualization:

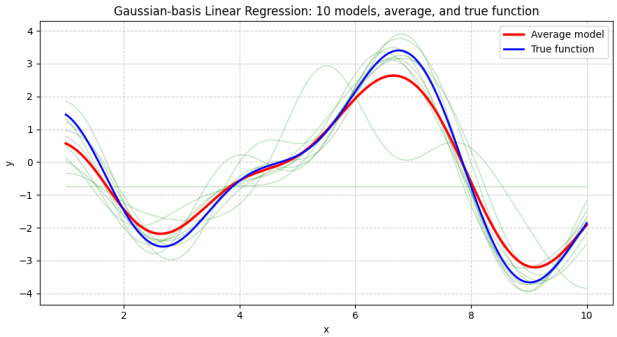

Figure 5: Ten fitted models, average model, and true function using Gaussian basis functions.

To analyze bias and variance, we repeated the entire fitting process from Task 1 a total of ten times, each time resampling 100 noisy data points from the same underlying distribution. For each run, we fit a linear regression model with the chosen number of Gaussian basis functions. The green curves correspond to the ten individual fits, the blue curve shows the true underlying function, and the red curve represents the mean of the fitted models.

When the number of basis functions is near the optimal complexity, the average model closely follows the true function while the individual green curves remain relatively concentrated, indicating low bias and moderate variance.

3.2 Train vs Test Errors:

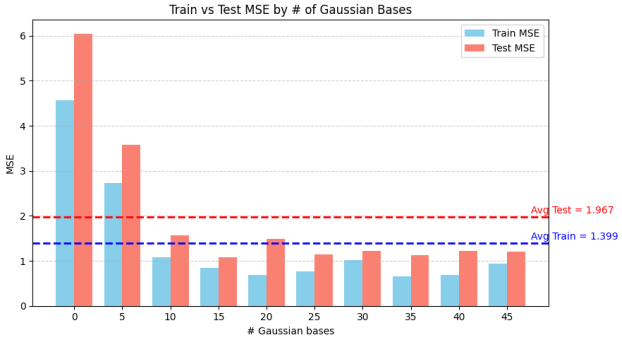

Figure 6: Average train and test MSE across 10 runs for different numbers of Gaussian bases.

For each model complexity D, we computed both the training and test mean squared errors (MSE) averaged over the 10 repetitions. The training error consistently decreases as D increases, indicating greater model flexibility. The test error initially drops with increasing D, reaching its minimum around intermediate complexity, then increases again for larger D values due to overfitting. The gap between train and test errors grows as D increases, which reflects increasing variance.

4 Task 3

4.1 Effect of Regularization Strength:

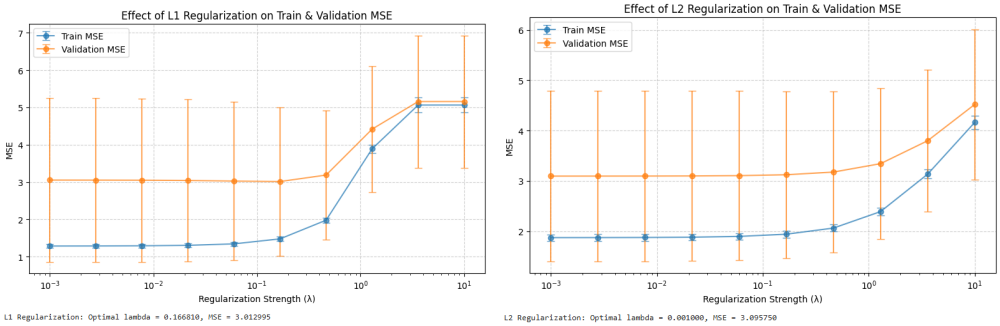

Figure 7: L1 & L2 regularization: train and validation MSE as a function of λ

We applied both L1 (Lasso) and L2 (Ridge) regularization on the model with $D = 45$ Gaussian bases, using 10-fold cross-validation to select the optimal regularization strength λ. For each λ value, the model was trained on 9 folds and validated on the remaining fold. The mean squared error (MSE) was then averaged across folds for both training and validation sets.

For L1 regularization, the minimum validation error was achieved at $\lambda \approx 0.1668$ with a validation MSE of approximately 3.0130. For L2 regularization, the optimal value was $\lambda \approx 0.0010$ with a validation MSE of 3.0958. As λ increases, the model becomes more regularized, leading to higher bias and lower variance, which stabilizes predictions but eventually causes underfitting when λ is too large. When λ is very small, the model has low bias but high variance and tends to overfit.

4.2 Bias-Variance Decomposition:

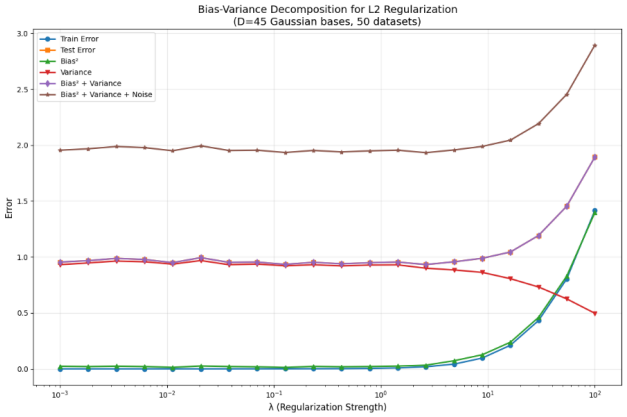

Figure 8: Bias–variance decomposition for L2 regularization across varying λ.

We decomposed the test error into bias, variance, and noise terms for L2 regularization across a range of λ values. As expected, the bias term grows steadily as λ increases, since strong regularization makes the model less flexible. Meanwhile, the variance term decreases with larger λ, reflecting increased stability of the model. The total error exhibits a U-shaped curve, with a minimum at $\lambda \approx 0.001$, where the tradeoff between bias and variance is optimal.

5 Task 4

5.1 Effect of L1 and L2 Regularization on Loss Landscapes:

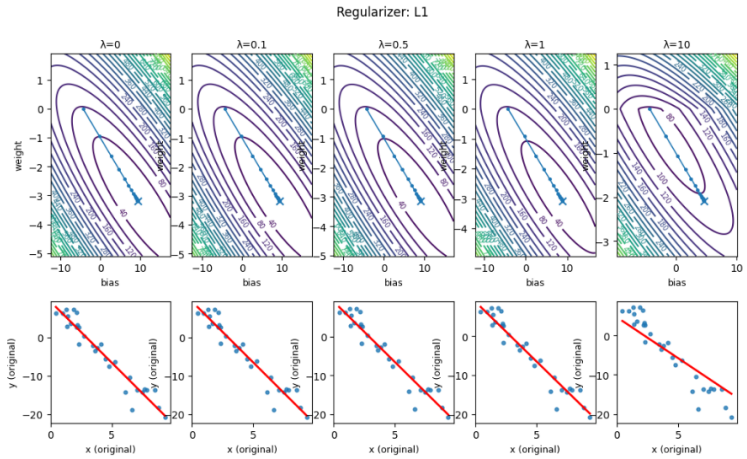

Figure 9: L1 Regularization: Effect of Varying λ on Loss Contours and Optimization Paths

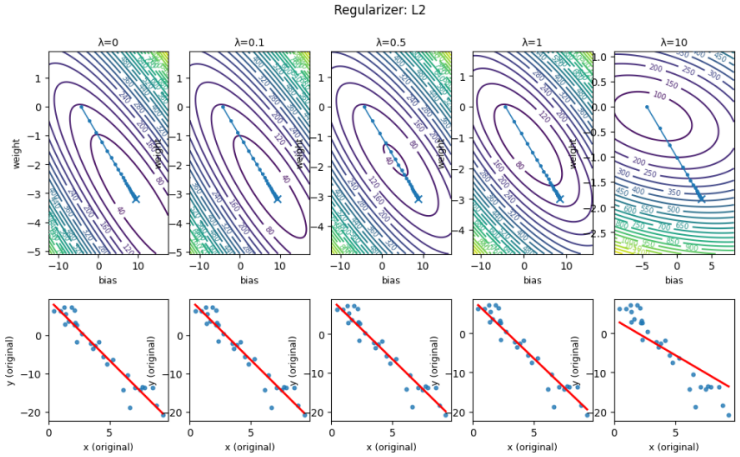

Figure 10: L2 Regularization: Effect of Varying λ on Loss Contours and Optimization Paths

As we increase the regularization strength, we observe a less good fit for both the Lasso (L1) and Ridge (L2) regularizations. This is due to the nature of what regularization does; L1 naturally attempts to make the weight vectors sparse while L2 naturally tries to reduce the overall weight vector magnitude to be as close to zero as possible. L1 achieves its goal by adding the L1 norm to the loss function, which is based off of the absolute value function, which is why we see sharp corners forming in the loss contours as we incrase the reg. strength. The sharp corners have large gradients which, through GD, brings specific components of the weight vector down to zero as fast as possible. Otherwise, L2 attempts a similar feat through the addition of the L2 norm to the loss function. This essentially adds circles centered at zero to the loss function, which essentially stretches and shifts the ellipses to be as close to zero-centered circles as possible. Thereafter, during training, we can observe that the minimum of the loss function shifts to being closer to the origin, which is why L2 regularization simply attempts to bring the overall weight values down as much as possible.

We can observe significant regularization effects when lambda = 10 for both cases. And, as for the other values, we can see that, as lambda increases, the loss surface starts to shift and stretch towards the origin, which is the overall goal of regularization to begin with.

6 Discussion and Conclusion

Through our testing, we have concluded that:

- Given a non-linear dataset, 5 standard Gaussian bases is generally a suitable number of bases to achieve a balance between model simplicity and validation error.
- The bias-variance tradeoff is clearly observed as we vary the number of Gaussian bases in our experiments. The "sweet spot" typically occurs around 10 bases, where the validation error reaches its minimum. Beyond this point, the validation error begins to rise, indicating overfitting. For smaller numbers of bases, both training and validation errors are high, indicating underfitting.
- Regularization strength plays a key role in preventing overfitting by reducing the model's complexity through weight shrinkage. We observe optimal λ values of 0.1668 for L1 and 0.0010 for L2 regularization, which yield the best validation results. These findings are supported by our bias-variance decomposition analysis, where the sweet spot occurs at similar λ values and the total validation error exhibits a U-shape curve.
- Regularization, whether L1 or L2, significantly affects the loss landscape. It shifts and stretches the loss contours, moving the center closer to the origin of the weight space. This illustrates how regularization influences the optimization process by guiding the weight vector toward a lower-magnitude region, promoting generalization and reducing overfitting.

7 Statement of Contributions

Part	**Contributor**
Part 1	Steven Thao
Part 2	Jawdat Al-Jabi
Part 3	Tal Smith
Part 4	Steven Thao
Code Review	All
Project Write-Up	All

Table 1: Project Contributions