Assignment 2 Comp 551

Angad Verma, Angela Moskal, Kai Turanski

October 24, 2024

Abstract

In this assignment, the effects of model complexity, regularization techniques, and cross-validation on linear regression models with non-linear basis functions were investigated. A linear regression model, transformed using Gaussian basis functions, was applied to explore the impact of increasing model complexity on performance, and to examine the transition from underfitting to overfitting. L1 and L2 regularization methods were implemented, and 10-fold cross-validation was employed to determine the optimal regularization strength. Additionally, loss contours were plotted to visualize the effects of regularization on the optimization landscape and the trajectories of gradient descent.

1 Linear Regression with Non-Linear Basis Functions

Data was sampled from the synthetic distribution $y(x) = \sin(\sqrt{x}) + \cos(x) + \sin(x) + \varepsilon$ where y is the target variable, x is the input data and ε is normally distributed noise with a variance of 1. The input was transformed with Gaussian bases, shown in Figure 1(a) and a linear regression model was fitted to the input, shown in Figure 1(b).

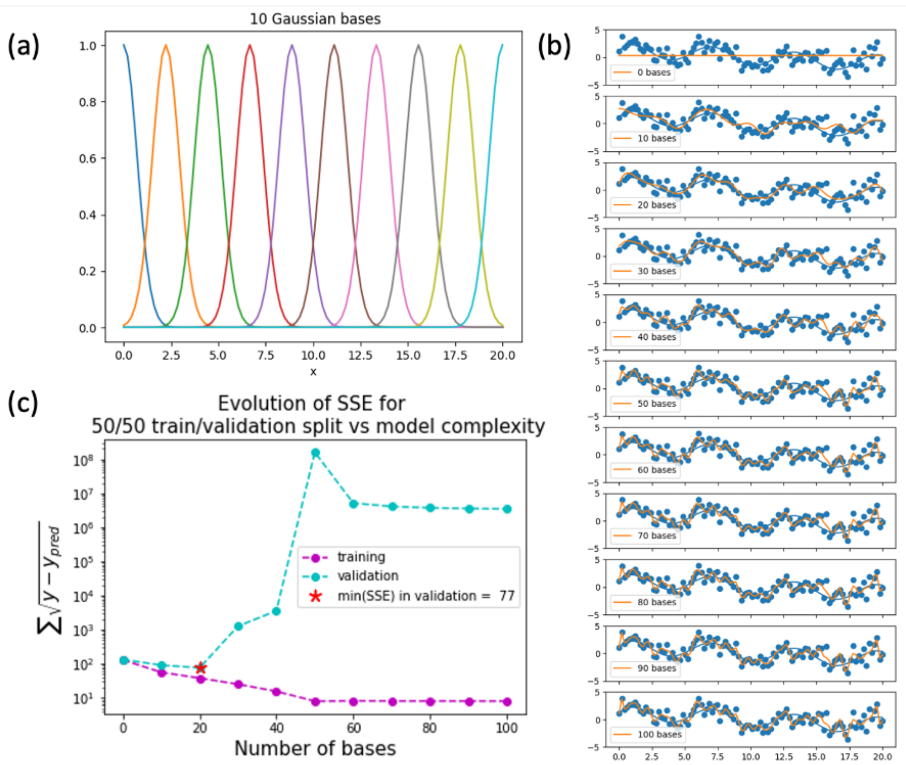

Figure 1: **a)** 10 Gaussian bases in the range of the synthetic dataset. **b)** Evolution of model fits with different number of bases. **c)** Evolution of training and test set SSE for models with different number of bases.

Figure 1(c) shows that the model performs best when there are 20 bases (with an MSE of 1). When the model has too few bases, it is underfitting and performs equally poorly on the training set and validation set. When the model has too many bases, it performs very well on the training set, but because it is overfitting, it performs poorly on the validation set. Minimizing the error in the validation set facilitated the selection of the best model since it shows which model is generalizable to all data from the distribution.

2 Bias-Variance Tradeoff with Multiple Fits

To validate the results from Task 1, the same experiment was repeated for 10 unique datasets, as shown in Figure 2(a).

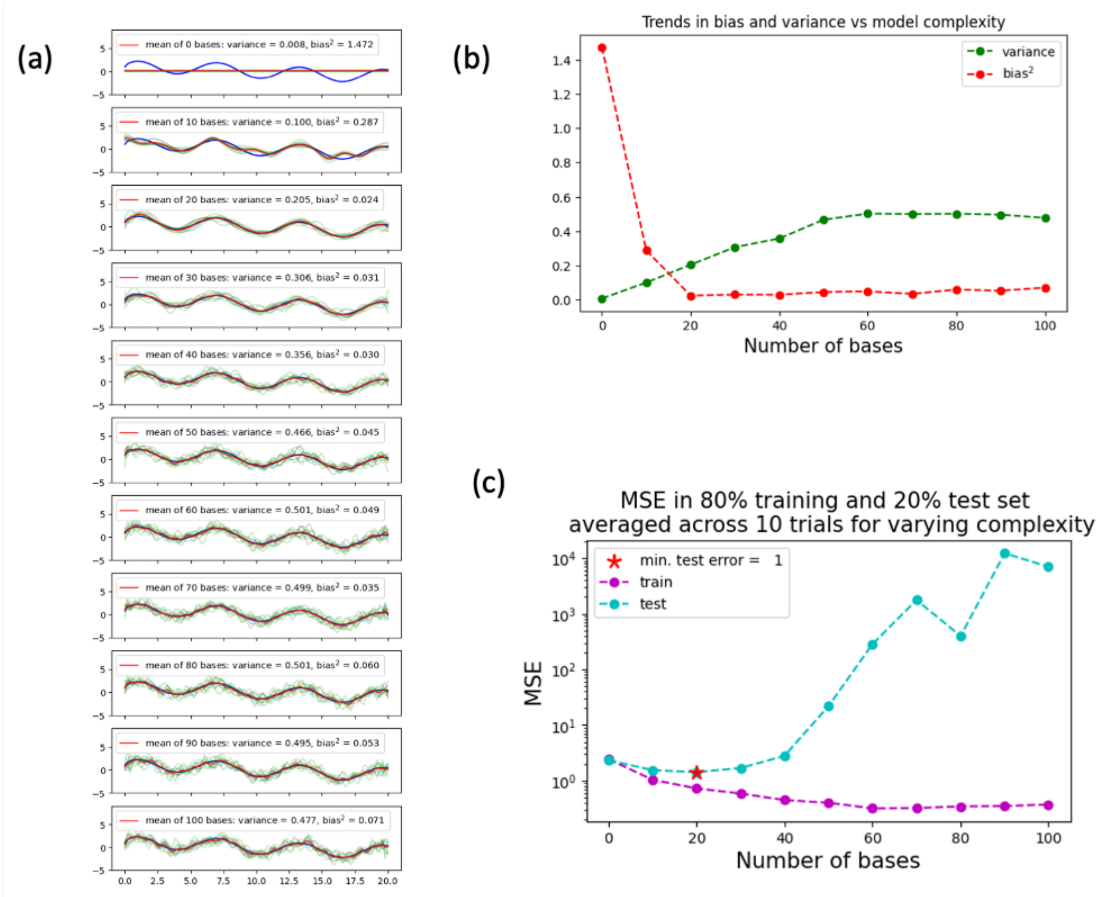

Figure 2: Evolution of models as complexity increases **a)** 10 models (green) were fitted to different datasets sampled from an underlying distribution (blue) and then averaged (red). **b)** Variation and bias trade-off as the number of bases in the model increases. **(c)** Training and test MSE as the number of bases in the model increases.

The bias/ variance plot in Figure 2(b) indicates that as the model's complexity increases, the variance increases. This is due to the fact that when the model exceeds 20 bases, it overfits. This is evidenced by Figure 2(c), in which the test mean squared error (MSE) steeply increasing at high model complexity, since each of the 10 individual models are tailored to the specific noise data they were trained on. Even though the complex models have more noise, low bias is still achievable by averaging over 10 models. Fewer than 20 bases and the model is too simple to fit to the noise or underlying function, so the variance is low but the bias is high. At around 20 bases is the sweet spot where the model is complex enough to fit to the underlying function, but not complex enough to fit the noise. This is the hyperparameter that best balances the bias-variance trade-off.

3 Regularization with Cross-Validation

Next L1 (lasso) and L2 (ridge) regression are introduced into our model and tuned to their best regularization strength (λ). This is implemented in a 10-fold cross-validation. Figure 3 shows how strong regularization increases the training MSE. However, by choosing an optimal λ, the validation MSE can be minimized. For L1 regularization, the optimal λ is between 0.001 and 0.008. For L2 regularization, the optimal λ is between 0.01 and 0.03.

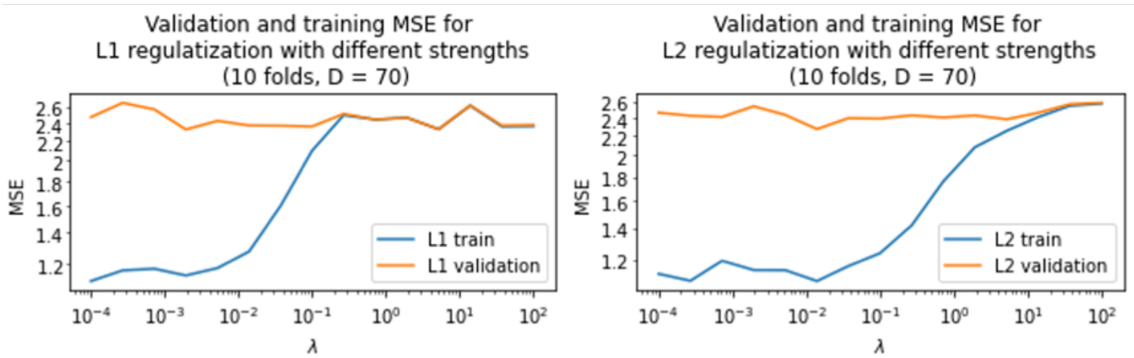

Figure 3: Effect of regularization strength (λ) in L1 and L2 regularization on the validation error and training error in 10-fold cross validation with 70 gaussian bases and averaging over 50 datasets.

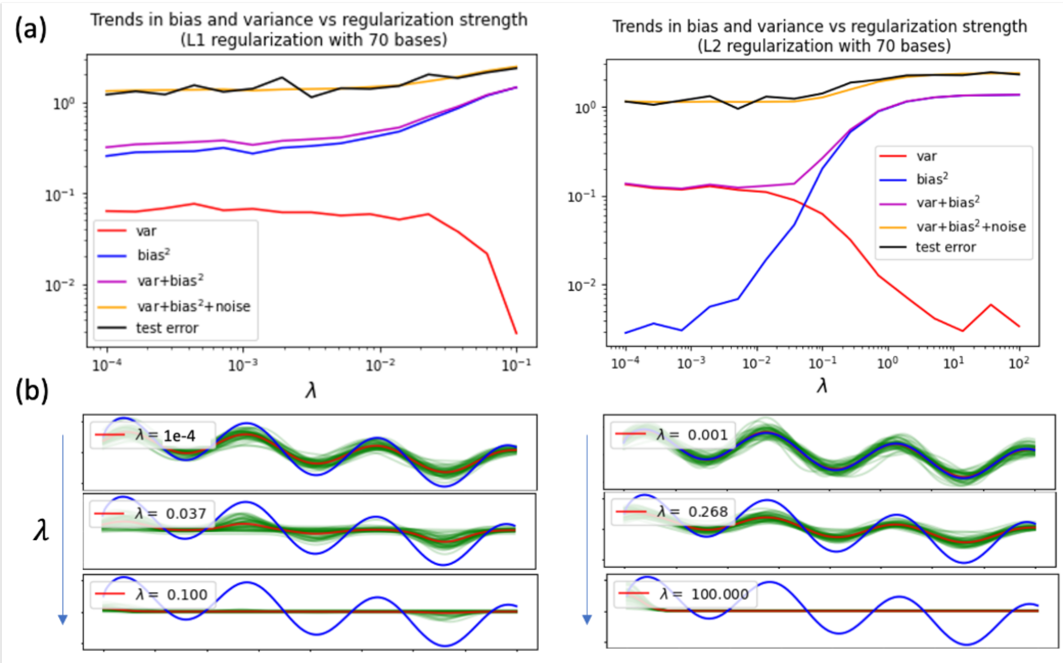

Figure 4: Effect of regularization strength (λ) in L1 and L2 regularization.**(a)** Bias-variance decomposition.**(b)** Extra: visualizing model fits for select regularization strengths.

Figure 4(a) clearly shows the decomposition of the test error into the bias, variance and noise. For strong regularization, the graph depicts a low variance but high bias. To better visualize the effects of regularization, the fits in Figure 4(b) were plotted to illustrate the different the mechanisms behind L1 and L2 regularization. These graphs clearly show how L1 works by eliminating some gaussian bases since some of the curves in the fit disappear while others remain. L2 on the other hand, does not eliminate certain gaussian bases.

4 Effect of L1 and L2 Regularization on Loss

In this task, the loss contours are plotted in order to visualize the paths taken by gradient descent in a linear regression model fitted to synthetic data drawn from the distribution $y = -4x + 10 + 2\varepsilon$.

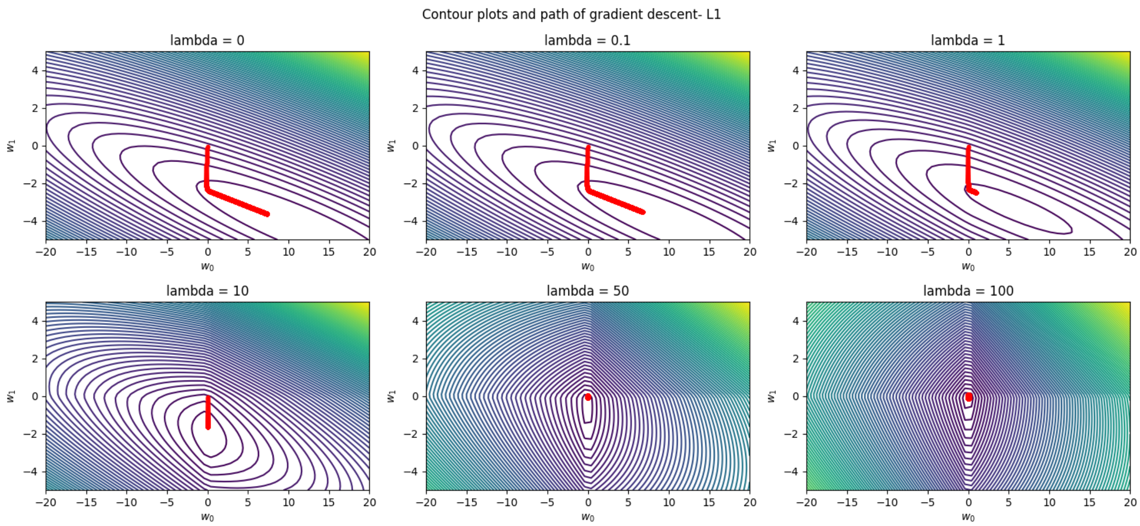

Figure 5: Effect of L1 regularization on the loss landscape for $\lambda = 0, 0.1, 1, 10, 50, 100$.

Figure 5 shows how L1 regularization encourages sparsity by pushing some weights towards zero. This is observable in the sharp changes in the descent path, which indicate that certain weights are being driven directly to zero, resulting in a sparser model. Further, the graphs show this regularization causes the contours to be more linear and the cost function to converge near zero, especially for large values of lambda.

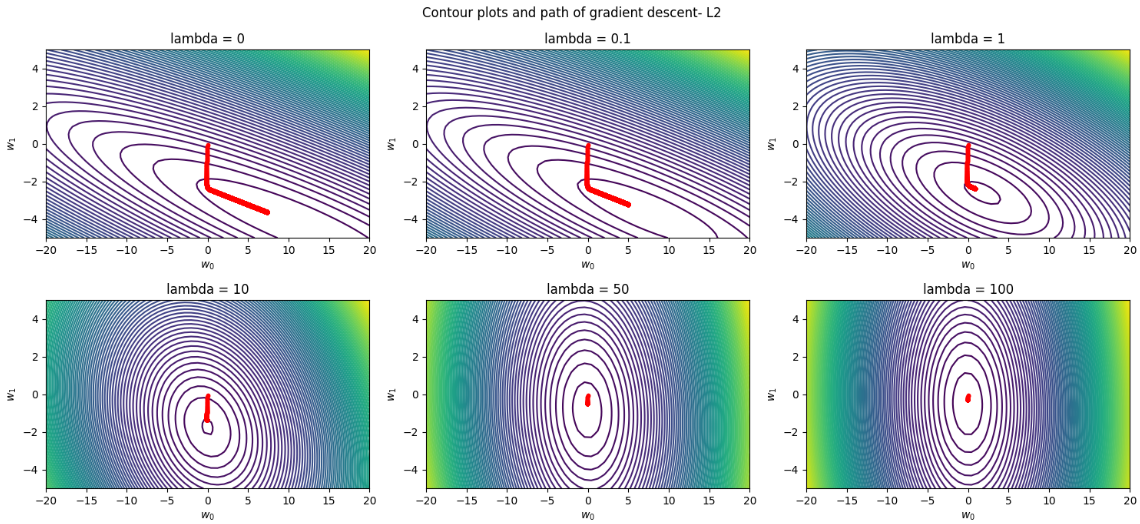

Figure 6: Effect of L2 regularization on the loss landscape for $\lambda = 0, 0.1, 1, 10, 50, 100$.

Figure 6 shows how L2 regularization penalizes large weights but has a decrease in sparsity compared to L1. L2 penalizes large weights more heavily, forcing the cost function to converge very close to zero. It also causes the contours to become more circular with increasing regularization strength, which helps in gradient descent.

5 Discussion and Further Improvements

The results of these four experiments demonstrate the need for balance between model complexity and generalization, and how techniques like regularization improve such balance. Tasks 1 and 2 show how an increasing number of non-linear bases increase the variance of the model, due to overfitting. The overfitting is especially apparent in the graphs which plot predicted values versus the dataset and ground truth, since the model increasingly favors sampled data values over the ground truth.
Tasks 3 and 4 demonstrate that it is possible to correct the overfitting bases of tasks 1 and 2 through the use of regularization. The error graphs of k-fold regularization show that validation error is decreased by nearly an order of magnitude with the right choice of λ. In the same vein, the bias/variance graphs show that stronger regularization decreases variance at the expense of higher bias, improving the model's generalizability. Lastly, the contour plots of varying regularization strengths illustrate exactly how extreme weights are limited through the application of regularization.
Overall, these experiments attest to the utility of regularization in improving generalization for complex models. In the future, it would be useful to experiment with the effects of regularization on increasingly complex models, like Adam and MLPs, and different types of non-linear bases.

Statement of contributions:

All team members independently completed the entire assignment, including data preprocessing, model implementation, and experiments. Afterward, they collaborated to compile the results, discuss key findings, and write the final report.