Bonus question:

4.1 Regularized Linear Regression

Figure 1 plots the variation in training/testing error with λ for regularized linear model.

- Best $\lambda = 10$:
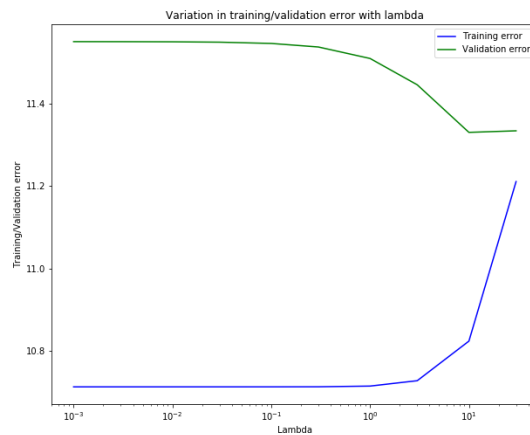
- training error: 11.605155

- test error: 12.6566689644

Figure 1: Variation in training/testing error with λ for regularized linear model

4.2 Selecting λ with quadratic features

Figure 2 plots the variation in training/testing error with λ for regularized linear model with quadratic features.

- Best $\lambda = 0.3$:

- training error: 4.256794

- test error: 4.82815820863

4.3 Selecting λ with cubic features

Figure 3 plots the variation in training/testing error with λ for regularized linear model with quadratic features.

- Best $\lambda = 3$:
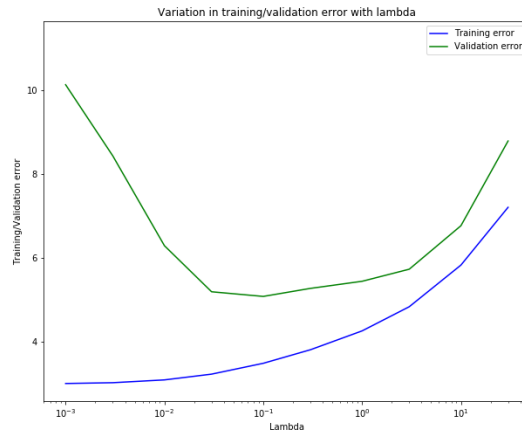
- training error: 3.645624

- test error: 4.73217608015

Figure 2: Variation in training/testing error with λ for regularized linear model with quadratic features

Summary

In above sections, we trained linear regression model with linear/quadratic/cubic features to predict Boston house price. We can observe that all three models we built embodied the power of regulation in that appropriate value of λ gives the "sweet region" — and even though the number of features grow dramatically as we square and even cube the features for more variation, the regulation term in the loss function successfully constrains the model from overfitting. The fact that models with linear, quadratic, cubic features are successively one better than the other shows that: given more features (variations) and proper control of overfitting, the more delicate model has better performance in prediction.
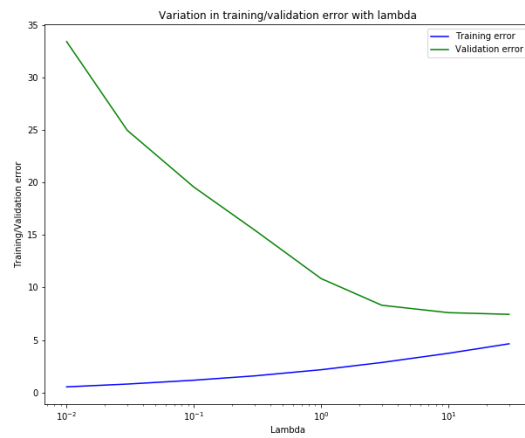
Figure 3: Variation in training/testing error with λ for regularized linear model with quadratic features