3.06 Regularization

Recap: What is overfitting?

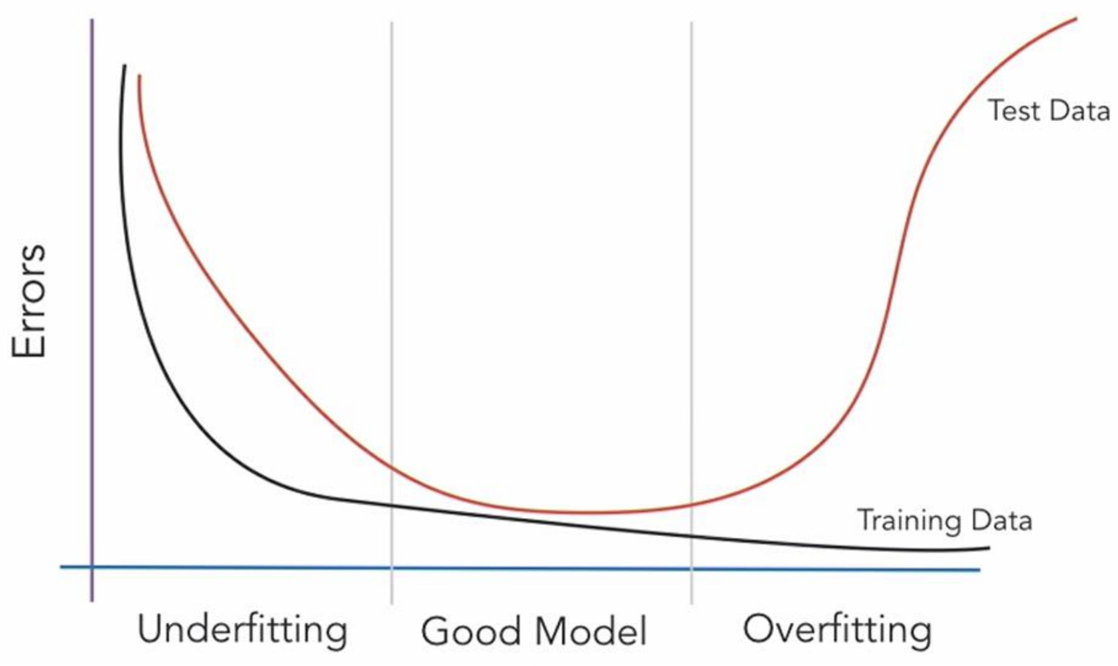
- When the model doesn’t pick up on the overall pattern of the data
- On the other hand, model emphasises specific data points too much

Regularization

- Technique used to reduce overfitting by discouraging overly complex models i.e. those with larger coefficients by forcing the models to adopt smaller coefficients

- Goal of regularization is to allow enough flexibility for the machine to learn the underlying patterns in the data but provide safeguards, so it doesn’t overfit

Regularization



Confucius ➔ Life is simple, but we insist on making it complicated

Occam's Razor Principle ➔ Whenever possible, choose the simplest answer to a problem

---

In our DS context, it means to choose the simplest model possible that will provide accurate predictions.

In other words, don't overcomplicate the model

Regularization models help to regulate the model so that the model does not get more complex than it really needs to be

Regularization – Models

- Ridge Regression & Lasso Regression
  - Adding a penalty to the loss function to constraint coefficients
  - Results in shrinking regression coefficients closer to 0 to make model simpler

Regularization – Ridge Regression

- We don't want our model to choose extremely large coefficients because that can lead to overfitting.
- Ridge regression helps us achieve this by adding a penalty to the model's cost function.
- It uses a technique called **L2 regularization**.
- The cost function of a linear regression model is as follows:

$$\text{Cost} = \text{RSS} = \sum_{i=1}^{n} \left(\text{y(i)} - \text{f(xi)}\right)^2$$

Regularization – Ridge Regression

- Our aim is to minimize the model's RSS since we want the difference between our true and predicted value to be as small as possible.
- In Ridge Regression, we include an additional parameter to the cost function, so it becomes:

$$\text{Cost} = \text{RSS} + \lambda * (sum\ of\ square\ of\ weights)$$

- We add the sum of square of model weights to the cost function.
- This means that the model's cost increases as it chooses larger weights (larger coefficients).
- This additional parameter acts as a constraint, and the model is now forced to choose smaller coefficients to minimise the cost function.

Regularization – Lasso Regression

- Lasso regression uses a technique called **L1 regularization**.
- **Like Ridge Regression**, it mitigates overfitting by adding a penalty to the cost function so that larger weights get penalized.
- **Unlike Ridge Regression**, instead of adding the sum of square of weights, Lasso adds the absolute value of weights to the cost.

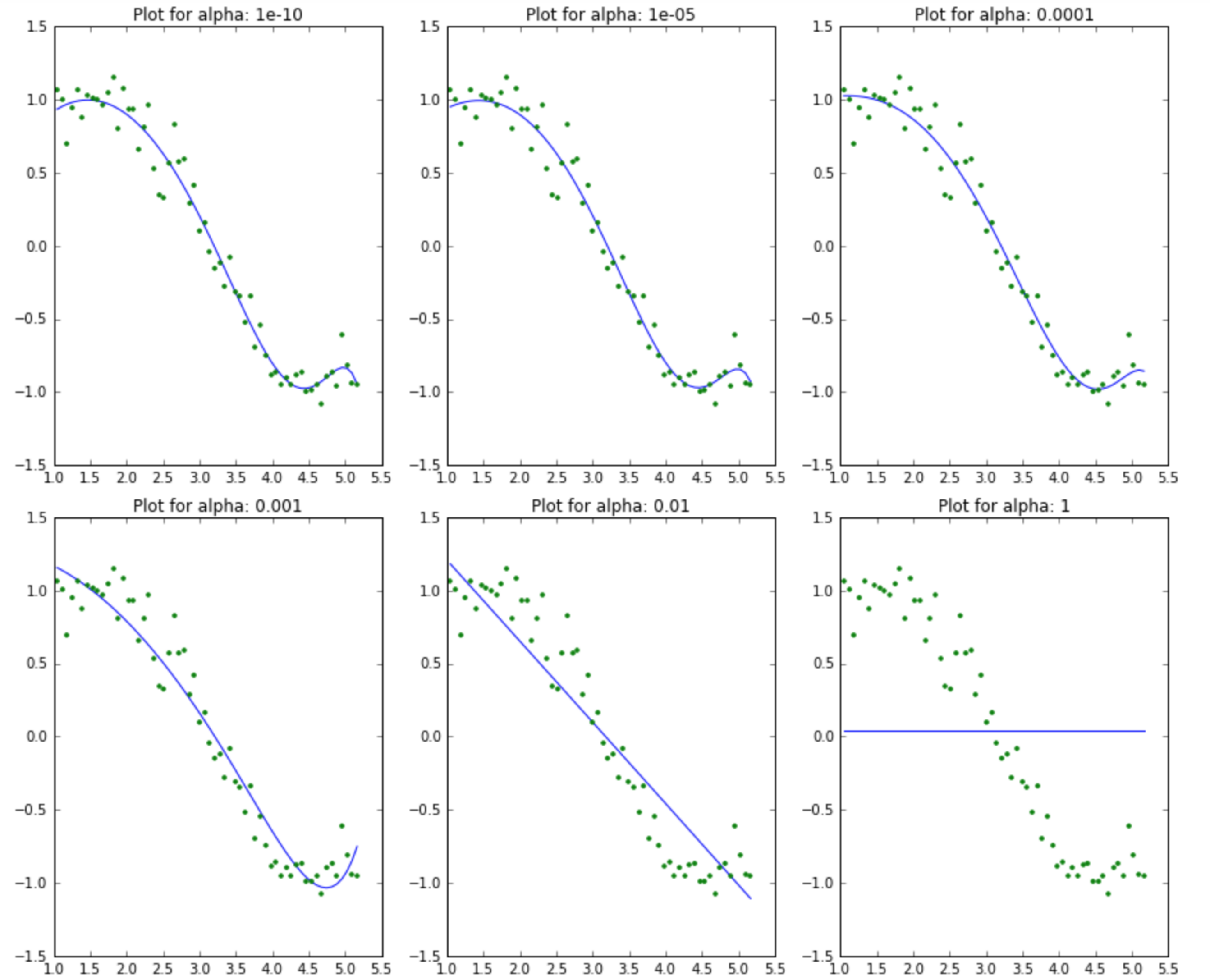
- The formula for lasso regression is:

$$\text{Cost} = \text{RSS} + \lambda * (sum\ of\ absolute\ weights)$$

Regularization – Alpha / Lambda

- A control in the model that limits how much weight to place on each feature included in the model
- The higher Alpha/Lambda, the less weight we can put on each feature
  - In other words, no one feature is going to be excessively influential in impacting the model results
- Finding the optimal control value requires experimentation

Regularization – Alpha / Lambda



Higher Alpha values can mitigate overfitting
e.g. Alpha = 0.0001 versus Alpha = 0.01

Regularization – Alpha / Lambda

- The cost becomes a lot higher as we increase the value of lambda

- The higher the Lambda, the lower the coefficients must be to minimise the cost function

- **Select a larger value for Lambda** if you want a model that generalizes better and heavily penalizes large coefficients