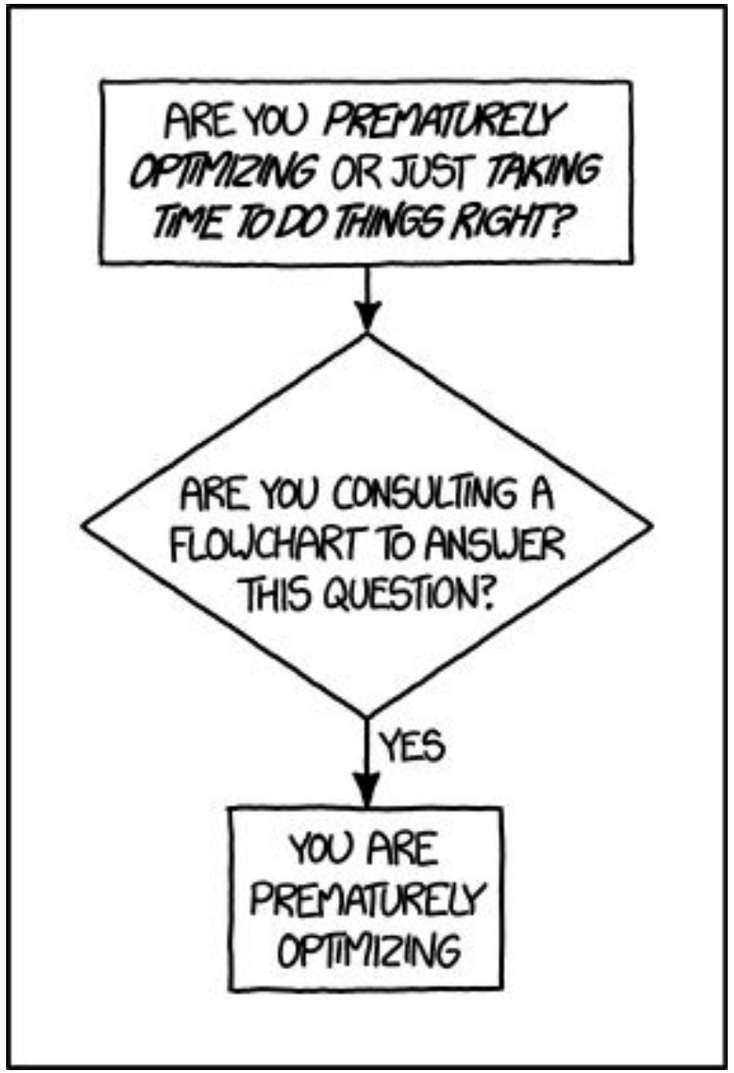
GENERAL ASSEMBLY

Regularization & Overfitting

Patrick D. Smith

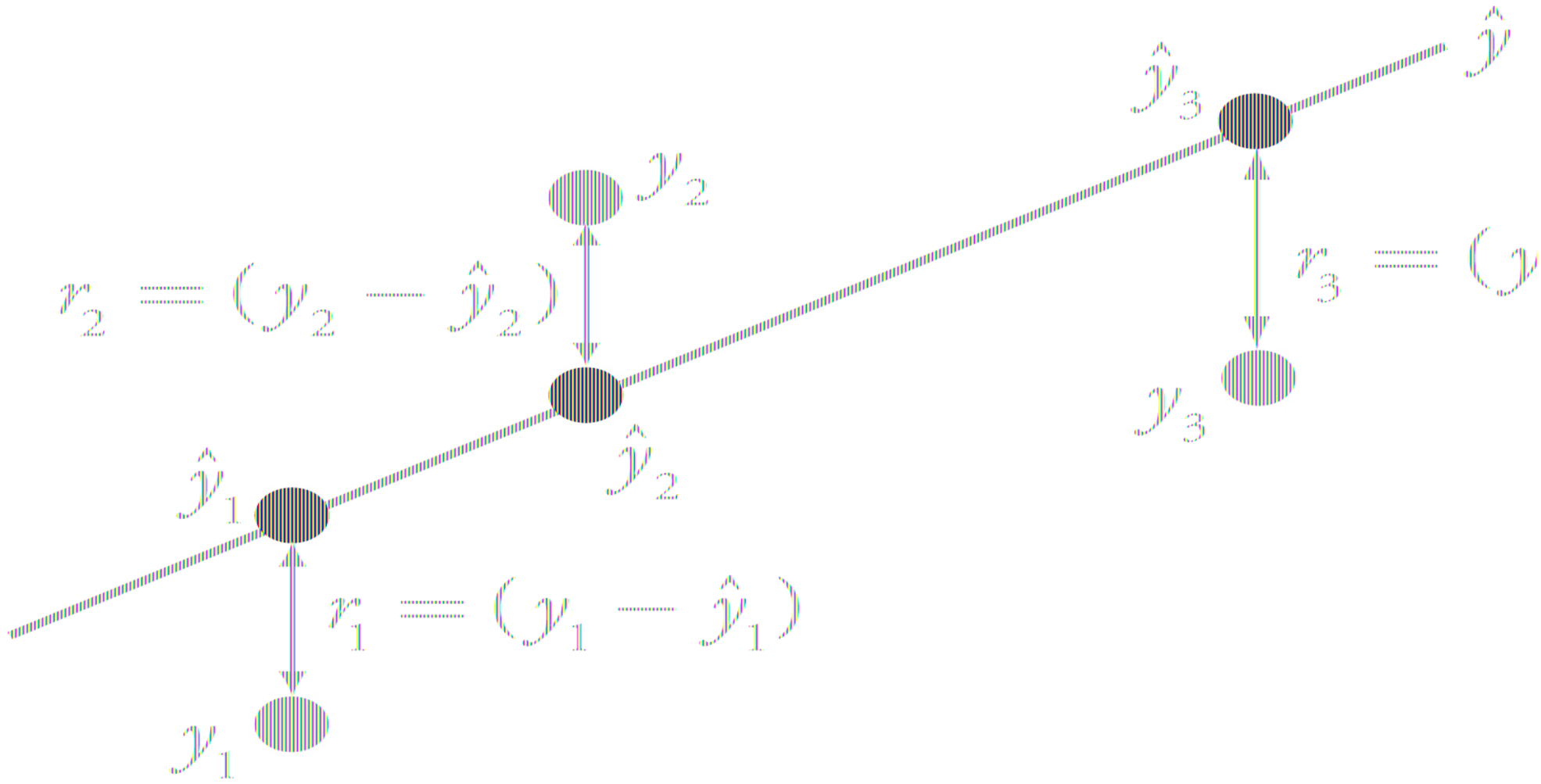
Lead Instructor, General Assembly DSI

Regularization and Overfitting



Regularization

Regularization and Overfitting



Regularization and Overtfitting

Ridge and Lasso Regression

Why would we want to regularize?

- Large enough to enhance the tendency of a model to overfit (as low as 10 variables might cause overfitting)
- Large enough to cause computational challenges. With modern systems, this situation might arise in case of millions or billions of features)

Regularization and Overtfitting

Ridge Regression (L2 Regularization)

- Adds a penalty to the square of the magnitude of the coefficients
- Effectively - “Punishes” large values by taking their square
- L2 seeks to have small values for all possible values of x
- L2 spreads error throughout the vector (dense matrix)

Regularization and Overtfitting

Objective = RSS + α * (sum of square of coefficients)

Here, α (alpha) is the parameter which balances the amount of emphasis given to minimizing RSS vs minimizing sum of square of coefficients. α can take various values:

1. **α = 0:**
    - The objective becomes same as simple linear regression.
    - We'll get the same coefficients as simple linear regression.
2. **α = ∞:**
    - The coefficients will be zero. Why? Because of infinite weightage on square of coefficients, anything less than zero will make the objective infinite.
3. **0 < α < ∞:**
    - The magnitude of α will decide the weightage given to different parts of objective.
    - The coefficients will be somewhere between 0 and ones for simple linear regression.

Regularization and Overtfitting

Lasso Regression (L1 Regularization)

- Adds a penalty equivalent to the absolute value of the magnitude of the coefficients
- L1 creates a sparse matrix - meaning that it is happy with very small values and very large values coexisting
- Lasso does variable selection automatically

Regularization and Overfitting

Objective = RSS + α * (sum of absolute value of coefficients)

Here, α (alpha) works similar to that of ridge and provides a trade-off between balancing RSS and magnitude of coefficients. Like that of ridge, α can take various values. Lets iterate it here briefly:

1. α = 0: Same coefficients as simple linear regression
2. α = ∞: All coefficients zero (same logic as before)
3. 0 < α < ∞: coefficients between 0 and that of simple linear regression

Regularization and Overtfitting