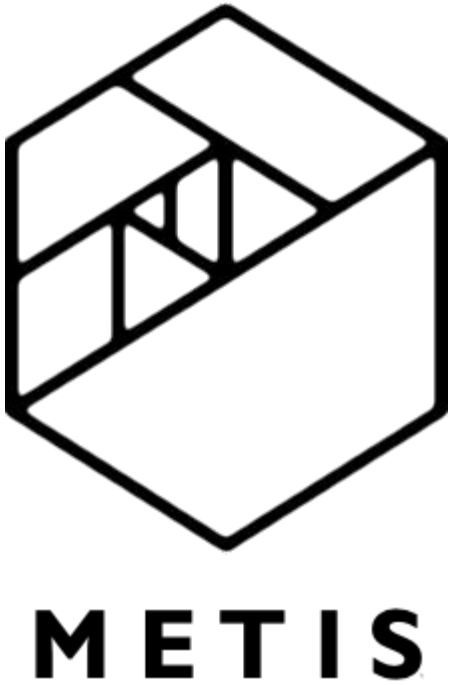


Regularization

What is Overfitting?

What is Overfitting?

1. Building a model that fits the training data too closely
2. Learning from noise in data, not the signal

What is Overfitting?

1. Building a model that fits the training data too closely
2. Learning from noise in data, not the signal.

What Causes Overfitting?

What is Overfitting?

1. Building a model that fits the training data too closely
2. Learning from noise in data, not the signal.

What Causes Overfitting?

1. Creating a model that is too complex
2. Evaluating the model on the same data used to train it

What is Overfitting?

1. Building a model that fits the training data too closely
2. Learning from noise in data, not the signal.

What Causes Overfitting?

1. Creating a model that is too complex
2. Evaluating the model on the same data used to train it

What Is the Impact?

What is Overfitting?

1. Building a model that fits the training data too closely
2. Learning from noise in data, not the signal.

What Causes Overfitting?

1. Creating a model that is too complex
2. Evaluating the model on the same data used to train it

What Is the Impact?

1. Model will not generalize well (not accurate on unseen data)
2. Low bias, high variance

How do you avoid overfitting?

1. MORE DATA (always a good answer, not always possible)
2. Feature engineering
3. Regularization!

Regularization

LASSO (l1) (Least Absolute Shrinkage and Selection Operator)

$$\sum_{i=1}^{n}\left(y_i-\beta_0-\sum_{j=1}^{p}\beta_j x_{ij}\right)^2+\lambda\sum_{j=1}^{p}|\beta_j| = \mathrm{RSS}+\lambda\sum_{j=1}^{p}|\beta_j|$$

Ridge (l2) (Long story about the name)

$$\sum_{i=1}^{n}\left(y_i-\beta_0-\sum_{j=1}^{p}\beta_j x_{ij}\right)^2+\lambda\sum_{j=1}^{p}\beta_j^2 = \mathrm{RSS}+\lambda\sum_{j=1}^{p}\beta_j^2$$

Regularization Intuition

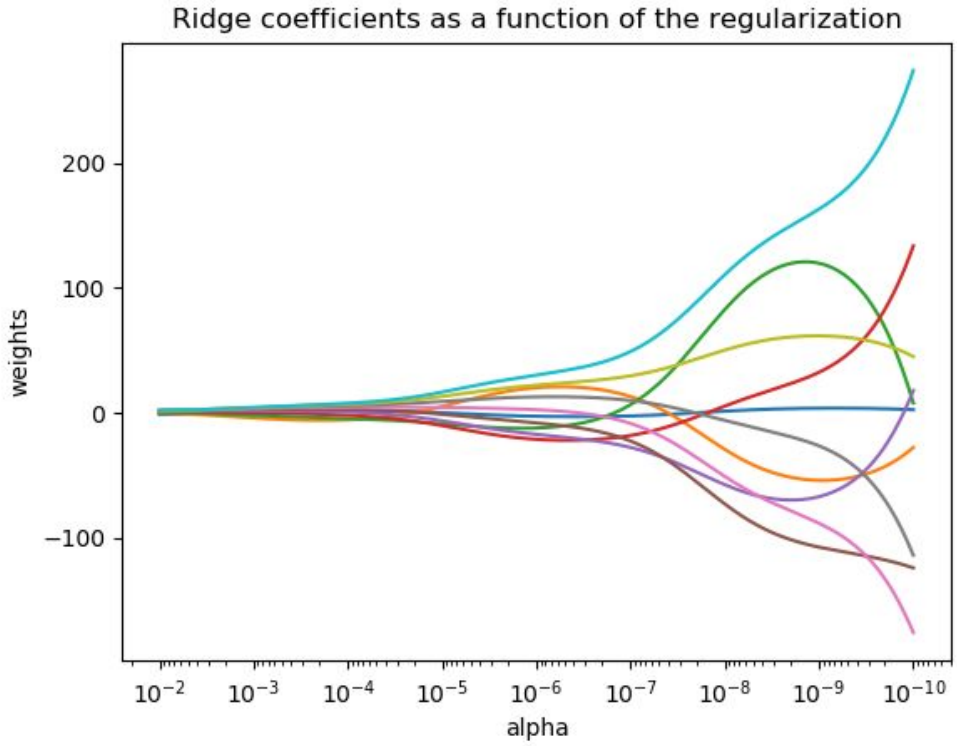
Ridge:

- If ƛ=0, normal RSS
- As ƛ→∞ coefficients will approach zero
- Shrinks all coefficients, still need to feature engineer

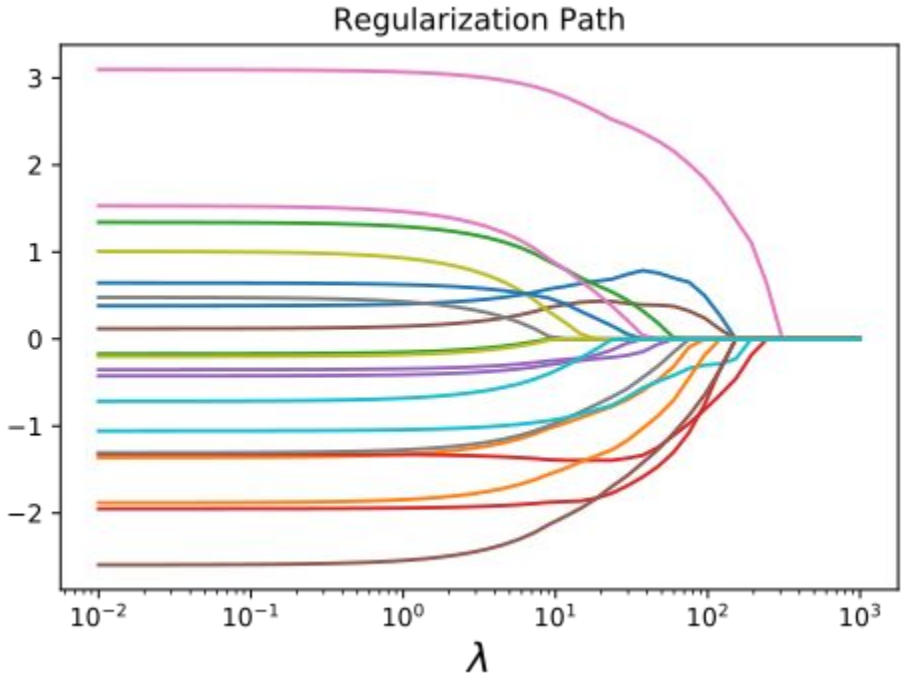
LASSO

- The absolute value will force to zero
- More interpretable

Ridge Visual Example



LASSO Visual Example



TO THE NOTEBOOK!!!!