

# Reflection Week 2

---

## Summary

In this week we reviewed and wandered deeper into Regularization concepts, including scaling, feature selections, sparsity, and a new LASSO regression, introduced as another technique besides Ridge Regression.

## Concepts

- Sparsity: a phenomenon in which there are many zeros or values that will not significantly impact a calculation. By identifying sparsity and get rid of the zero coefficients after applying regularization, we can reduce the number of features and simplify our models to better generalize trends.
- Steps on choosing  $\lambda$  for regularization:
  1. Train a model using gradient descent, applying either LASSO or Ridge Expression with multiple tries using different values of  $\lambda$
  2. Compute validation error of these models
  3. Track the  $\lambda$  with the smallest validation error
  4. Return this value  $\lambda^*$  and estimated future error  $RSS_{test}(\hat{w}_{ridge(\lambda^*)})$
- Feature selection for regularization: Start with a full model, then "shrink" the coefficients near 0. Non-zero coefficients above a certain threshold  $\alpha$  would be considered selected as important.
- LASSO regression:  $\hat{w} = \min(RSS(W) + \lambda ||w||_1)$

## Concerns

- How to interpret the visualization of LASSO vs. Ridge solutions? What do the eclipses represent, and why do they have that specific shape?
- How to determine the  $\alpha$  value in choosing features?
- Is there an evaluation method that combines both choosing multiple features and allowing polynomials/exponents of specific ones, such as  $\hat{y} = \hat{w}_0 + \hat{w}_1 h[1] + \hat{w}_2 h[1]^2 + \hat{w}_3 h[2] + \hat{w}_4 \log(h[3])$ ?