

Crib 2

# 02 Ridge Regression

by Alvin Wan . [alvinwan.com/cs189/fa17](http://alvinwan.com/cs189/fa17)

## 1 Ridge Regression

1. objective:  $\min_w \|Xw - y\|_2^2 + \lambda\|w\|_2^2$
2. solution:  $w^* = (X^T X + \lambda I)^{-1} X^T y$
3. Why regularize?
  - prevents overfitting (models that are overly complex, too large etc.)
  - motivated by Gaussians (next week)
  - makes a number of assumptions about our data (see quiz 2, problem 2)
4. Why L2-norm? Rotationally invariant (doesn't change under rigid transformations—rotations or reflections)

## 2 Data and Overfitting

1. Split data into train, validation, and test (evaluate only once)
2. Expressive models can *overfit*—fit to noise in data.
3. Use regularization term to prevent overfitting.
4. Use k-fold cross-validation to tune hyperparameters.
5. Validation error is not a direct proxy for generalization error (irreducible error, introduced later)