



# Regularization/Feature Scaling

10/20/2021

# Lesson Plan

- Regularization
- Lasso (L1)
- Ridge (L2)
- Regularization Parameter
- Feature Scaling
- Normalization
- Standardization

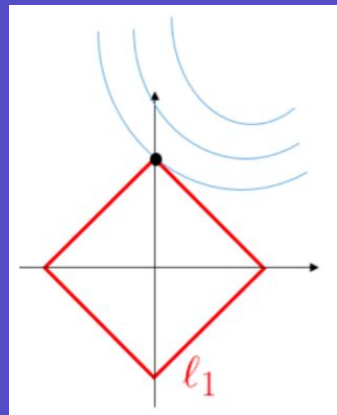
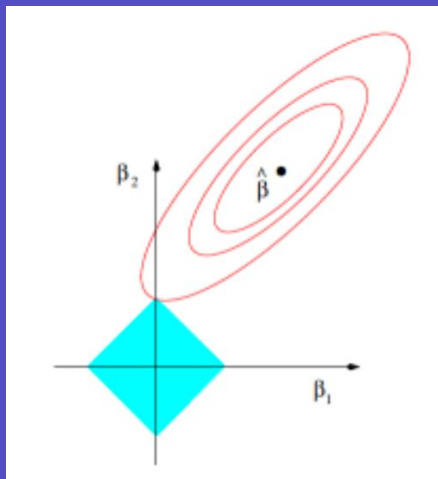
# Regularization

- Definition: used to tune the model by adding an additional term to the cost function
- Penalizes certain weights/parameters that the model is learning
- Goal: reduce variance in model (prevent overfitting)
- Lasso/Ridge regularized regression

# Lasso (L1)

- Least absolute shrinkage and selection operator
- Adds a regularization parameter multiplied by the absolute sum of the parameters
- Have to minimize both the RSS (residual sum of squares) and regularization term
- Can think of the regularization term as a constraint: prevents the parameters from taking large values
- Goal: find the spot that minimizes the sum of RSS and regularization term

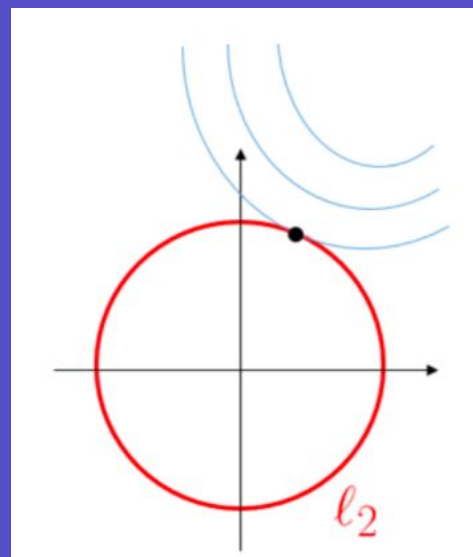
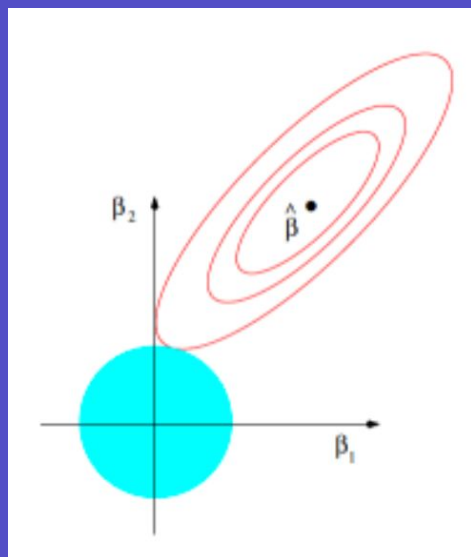
$$\sum_{i=1}^n \left( y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^p |\beta_j| = \text{RSS} + \lambda \sum_{j=1}^p |\beta_j|.$$



# Ridge (L2)

- Adds a regularization parameter multiplied by the squared sum of the parameters
- Have to minimize both the RSS (residual sum of squares) and regularization term

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



# Regularization Parameter

- Low value: high variance/low bias (overfit)
- High value: low variance/high bias (underfit)



# Feature Scaling

- An important step in preprocessing
- A lot of algorithms are biased towards numerically larger values
- Feature scaling prevents certain large features from impacting the model

# Normalization

- Often refers to min-max scaling
- $X_{\text{new}} = (X - X_{\text{min}}) / (X_{\text{max}} - X_{\text{min}})$
- Data scaled to range [0, 1]
- Heavily affected by outliers

# Standardization

- Also known as z-score normalization
- $X_{\text{new}} = (X - \text{mean}) / \text{Std}$
- Center of data becomes 0
- Not affected by outliers (not bounded to a specific range)