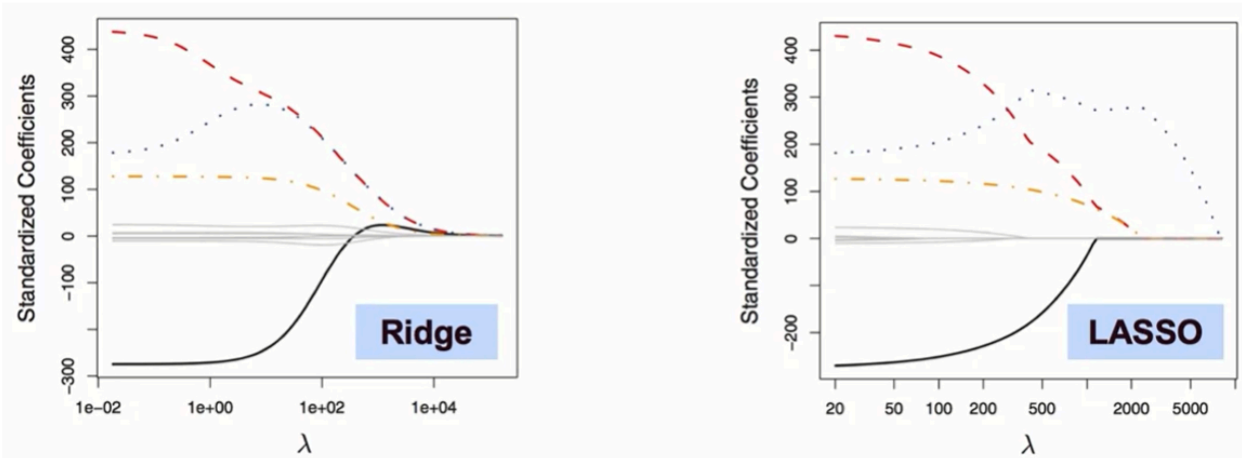
SUPERVISED MACHINE LEARNING: REGRESSION

MODULE 5:

REGULARIZATION DETAILS

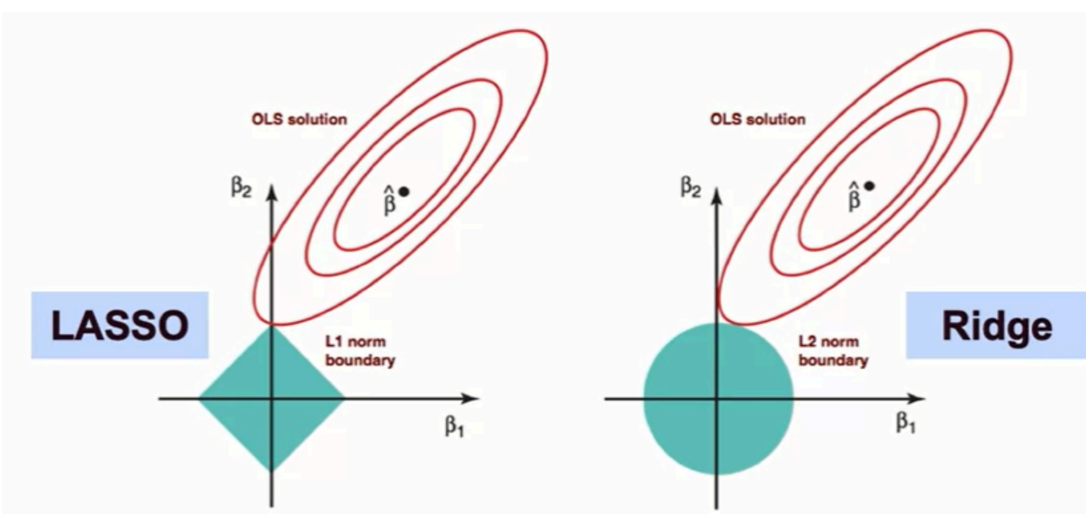
TABLE OF CONTENTS

Analytical View



- **Idea:** Regularization forces coefficients to be smaller, shrinking their possible range.
- **Effect:**
  - Smaller coefficients → simpler model → lower variance.
  - Large coefficients → high sensitivity → high variance.
- **Intuition:** Reducing coefficient magnitude limits how strongly features affect predictions, stabilizing the model.

Geometric View



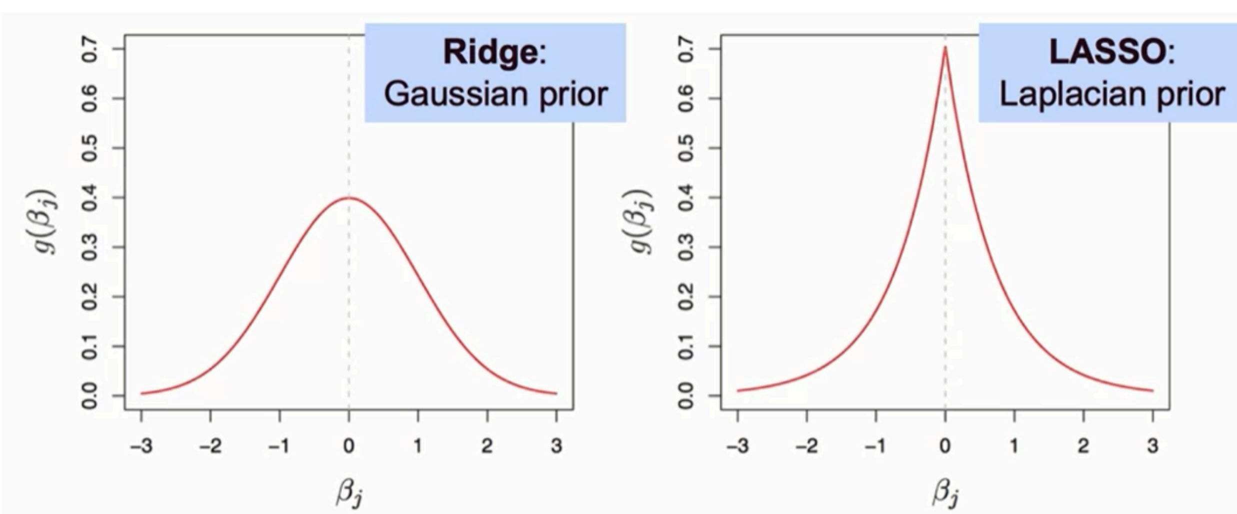
- **Optimization objectives of Ridge/LASSO:**

Ridge $$\underset{\beta}{\text{minimize}} \left\{ \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 \right\} \quad \text{subject to} \quad \sum_{j=1}^{p} \beta_j^2 \le s.$$

LASSO $$\underset{\beta}{\text{minimize}} \left\{ \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 \right\} \quad \text{subject to} \quad \sum_{j=1}^{p} |\beta_j| \le s$$

- **Optimization Perspective:**
  Regularization constrains the coefficient space by limiting the sum of squared (Ridge) or absolute (LASSO) coefficient values.
- **Ridge (L2):** Constraint forms a **circle** — intersection with the OLS contour can occur anywhere → coefficients are *shrunk but not zeroed*.
- **LASSO (L1):** Constraint forms a **diamond** — intersection often touches corners → some coefficients become *exactly zero*.
- **Result:** The diamond shape of LASSO explains its ability to perform **feature selection**, unlike Ridge.

Probabilistic View



- **Bayesian Interpretation:**
  Regularization imposes **prior distributions** on model coefficients.
  - **Ridge (L2):** Assumes a **Gaussian (normal) prior** — coefficients likely near zero but continuous.
  - **LASSO (L1):** Assumes a **Laplacian prior** — sharper peak at zero → more coefficients exactly zero.
- **Lambda (λ):** Controls the variance of these priors.
  - Higher λ → smaller variance → stronger belief that coefficients ≈ 0.
- **Goal:** Balance bias and variance by penalizing large coefficients, yielding stable and generalizable models.

Key Takeaways

- Regularization reduces **variance** at the cost of a small **bias increase**.
- Ridge and LASSO offer trade-offs between smooth shrinkage and feature elimination.
- The **analytic**, **geometric**, and **probabilistic** views reveal *how and why* regularization simplifies models and improves generalization.