📘 Pre-Session Notes

Topic: Regularization – The Secret to Taming Overfitting

🔍 1. Introduction to Regularization

What is Overfitting?
Overfitting occurs when a model performs well on training data but poorly on unseen test data. This happens because the model learns the noise in the training data instead of the general patterns.

Regularization – The Cure:
Regularization is a technique used to reduce overfitting by discouraging complex models. It does this by penalizing large coefficients in the model's loss function.

Key Idea:
Regularization adds a penalty term to the original cost function (like Mean Squared Error) to restrict the magnitude of model parameters (weights).

> New Cost = Original Loss + Penalty (Regularization term)

🧮 2. Lasso (L1 Regularization)

Definition:
Lasso (Least Absolute Shrinkage and Selection Operator) adds a penalty equal to the absolute value of the coefficients (L1 norm).

Cost Function:

$$J(\theta) = \text{MSE} + \lambda \sum_{j=1}^{n} |\theta_j|$$

Key Characteristics:

- **Encourages sparsity: Some coefficients become exactly zero.**
- **Useful for feature selection.**
- **Tends to create simpler models by eliminating unimportant features.**

Example Use Case:
Useful when you believe many features are irrelevant and should be ignored.

🧲 3. Ridge (L2 Regularization)

Definition:
Ridge Regression adds a penalty equal to the square of the magnitude of the coefficients (L2 norm).

Cost Function:

$$J(\theta) = \text{MSE} + \lambda \sum_{j=1}^{n} \theta_j^2$$

Key Characteristics:

- **Shrinks coefficients toward zero, but never exactly zero.**
- **Keeps all features, but reduces their impact.**
- **Stabilizes the model by avoiding large weights.**

Example Use Case:
Effective when you suspect all features are useful but want to prevent overfitting.

🧠 4. Polynomial Regression and Regularization

Why it Matters:
Polynomial regression increases model complexity by adding higher-degree terms (e.g., x2x^2, x3x^3). While it may improve training accuracy, it is highly prone to overfitting.

Solution:
Apply L1 or L2 regularization to polynomial regression to:

- Control coefficient growth of high-degree terms.
- Keep the model generalizable to test data.

Example:
A degree-10 polynomial may fit the training set perfectly but can oscillate wildly on test data. Regularization helps tame this.

⚖️ 5. Choosing the Right Regularization Type

Criteria	Lasso (L1)	Ridge (L2)
Feature selection	✅ Good (sets some weights to 0)	❌ Keeps all features
Works well when	Many features are irrelevant	All features are relevant
Solution	Sparse	Smooth
Optimization	Less stable due to absolute value	Stable (differentiable)

Elastic Net:

A combination of L1 and L2:

J(θ)=MSE+λ1∑∣θj∣+λ2∑θj2J(\theta) = \text{MSE} + \lambda_1 \sum |\theta_j| + \lambda_2 \sum \theta_j^2

Use when you want both sparsity and regularization stability.

📌 Summary

- **Regularization is essential to prevent overfitting.**
- **L1 (Lasso) leads to simpler models by feature elimination.**
- **L2 (Ridge) retains all features but keeps them in check.**
- **Regularization is especially important when using complex models, like polynomial regression.**
- **Choice depends on your dataset characteristics and goals.**
