

Lasso and Ridge Regression



Act like my Machine Learning teacher and teach me Lasso Regression and a complete guide and maths intuition and 25 mcq questions and answers

Absolutely! Let's dive into **Lasso Regression**, step by step. I'll structure the lesson like this:

What You'll Learn:

1. Introduction to Lasso Regression
2. Need for Regularization
3. Mathematical Intuition
4. Lasso vs Ridge vs Linear Regression
5. Geometric View of Lasso
6. Advantages & Disadvantages
7. Python Implementation
8. 25 MCQs with Answers

1. Introduction to Lasso Regression

Lasso (Least Absolute Shrinkage and Selection Operator) is a type of **linear regression** that includes **L1 regularization**. It helps prevent **overfitting** and can also **shrink some coefficients to exactly zero**, effectively performing **feature selection**.

2. Why Regularization?

When you train a model with **many features**, there's a chance it will fit too closely to the training data (**overfit**) and perform poorly on new data.

Regularization:

- Adds a **penalty term** to the loss function.
- Forces the model to keep weights (coefficients) small.



3. Mathematical Intuition

Standard Linear Regression:

Loss Function:

$$RSS = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij})^2$$

Lasso Regression adds L1 penalty:

$$Loss = \underbrace{\sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij})^2}_{\text{Residual Sum of Squares (RSS)}} + \underbrace{\lambda \sum_{j=1}^p |\beta_j|}_{\text{L1 Penalty}}$$

Where:

- $\lambda \geq 0$ is the **regularization strength**.
- Higher λ means more penalty (more shrinkage of coefficients).



4. Lasso vs Ridge vs Linear

Feature	Linear Regression	Ridge (L2)	Lasso (L1)
Penalty Term	None	$\lambda \sum \beta_j^2$	$(\lambda \sum \beta_j)$
Coefficients Shrink?	No	Yes	Yes (some to zero)
Feature Selection	✗	✗	✓



5. Geometric Intuition

- Lasso uses an **L1-norm constraint**, which forms a **diamond-shaped region** in the coefficient space.
- Because of the sharp corners of the diamond, the optimization often "sticks" at corners → some coefficients = 0.



6. Advantages & Disadvantages



Advantages:

- Feature selection
- Reduces overfitting
- Simple and interpretable



Disadvantages:

- Can be unstable when features are highly correlated
- Doesn't perform group selection (only picks one of many correlated variables)

7. Lasso in Python

```
python

from sklearn.linear_model import Lasso
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_squared_error

# Sample data
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2)

# Apply Lasso
lasso = Lasso(alpha=1.0)
lasso.fit(X_train, y_train)

# Predict
y_pred = lasso.predict(X_test)

# Evaluate
print("MSE:", mean_squared_error(y_test, y_pred))
print("Coefficients:", lasso.coef_)
```

8. 25 MCQs on Lasso Regression

Q1. What is the penalty used in Lasso Regression?

- a) L2
- b) L1
- c) L0
- d) None

Ans: b)

Q2. Lasso can shrink some coefficients to:

- a) Zero
- b) Infinity
- c) Negative values only
- d) Mean of data

Ans: a)

Q3. In Lasso regression, increasing λ leads to:

- a) Larger coefficients
- b) Smaller coefficients
- c) No change
- d) Overfitting

Ans: b)

Q4. Lasso helps in:

- a) Model selection
- b) Feature selection
- c) Clustering
- d) Normalization

Ans: b)

Q5. Lasso regression reduces:

- a) Bias
- b) Variance
- c) Both bias and variance
- d) None

Ans: b)

Q6. Which norm is used in Lasso?

- a) L1
- b) L2
- c) Frobenius
- d) Manhattan

Ans: a)

Q7. When $\lambda = 0$, Lasso becomes:

- a) Ridge
- b) Linear Regression
- c) Logistic Regression
- d) SVM

Ans: b)

Q8. Lasso prefers:

- a) All features equally
- b) Dense models
- c) Sparse models
- d) Random features

Ans: c)

Q9. Lasso works well when:

- a) Many correlated features
- b) Few non-zero features
- c) All features are relevant
- d) Data is non-linear

Ans: b)

Q10. The main difference between Lasso and Ridge is:

- a) Type of data used
- b) Loss function
- c) Type of penalty
- d) Use of cross-validation

Ans: c)

Q11. Lasso is best for:

- a) Dimensionality reduction
- b) Time series
- c) Classification
- d) Feature extraction

Ans: a)

Q12. What happens to irrelevant features in Lasso?

- a) Grows large
- b) Becomes NaN
- c) Shrunk to zero
- d) Stays constant

Ans: c)

Q13. Lasso's objective function is:

$$\min_{\beta} \sum (y - X\beta)^2 + \lambda \sum |\beta|$$

True / False?

Ans: True

Q14. Does Lasso perform variable selection?

Ans: Yes

Q15. Does Ridge perform variable selection?

Ans: No

Q16. Lasso penalty is:

- a) Convex
- b) Non-convex
- c) Linear

d) Non-differentiable

Ans: a) (but also note: L1 is convex but not differentiable at 0)

Q17. Which of the following can be zero in Lasso?

- a) Error
- b) Intercept
- c) Coefficients
- d) X-values

Ans: c)

Q18. What's the typical result of Lasso on high-dimensional data?

- a) Dense models
- b) Models with all features
- c) Models with few selected features
- d) Logistic models

Ans: c)

Q19. In Lasso, what does a very high lambda do?

- a) Fit well
- b) Selects all features
- c) Makes model too simple
- d) Enhances accuracy

Ans: c)

Q20. Lasso regression is primarily used for:

- a) Classification
- b) Regression with irrelevant features
- c) Clustering
- d) PCA

Ans: b)

Q21. Lasso can't handle:

- a) Sparse data
- b) Multicollinearity well
- c) Linear relationships
- d) Overfitting

Ans: b)

Q22. What's the role of alpha in Lasso in sklearn?

- a) It's learning rate
- b) It's max iterations
- c) It's the regularization strength
- d) None

Ans: c)

Q23. Lasso selects features by:

- a) Correlation
- b) Randomness
- c) Shrinking less useful coefficients to zero
- d) Clustering

Ans: c)

Q24. If we use StandardScaler with Lasso, it is to:

- a) Normalize target
- b) Scale inputs
- c) Increase speed
- d) Nothing

Ans: b)

Q25. Which regression shrinks some coefficients to zero?

- a) Ridge
- b) OLS
- c) Lasso
- d) None

Ans: c)

Would you like:

- A PDF of this guide?
- A small dataset to try Lasso yourself?
- A notebook to test the MCQs interactively?

Let me know how you'd like to proceed.