Editorial-W2A1: Regression Models and Regularization Fundamentals

Question 1

A marketing analyst is working on a dataset containing customer demographics and purchasing history to predict future spending behavior. The dataset contains many features, but the analyst suspects that some of them are irrelevant to the prediction. They want to build a model that can automatically exclude irrelevant features during training.

Which regularization technique should they use?

Options:
A) Ridge Regression
B) Lasso Regression
C) Linear Regression
D) Logistic Regression

Answer: B
Explanation: Lasso regression performs both regularization and feature selection by shrinking some coefficients to exactly zero, effectively removing irrelevant features.

Question 2

A data science team is evaluating three regression models (Model 1, Model 2, Model 3) trained on a dataset with five features. The weight vectors for each model are provided below:

- **Model 1:** [0.6, 0, 0, 0.2, -0.34]
- **Model 2:** [0.9, 0.24, -0.23, 0.45, -0.61]
- **Model 3:** [0.23, 0.05, -0.06, 0.12, -0.19]

The team used **Linear Regression**, **Ridge Regression**, and **Lasso Regression** but forgot to label the models. Based on the weight vectors, identify which model corresponds to each method.

Options:
A) Model 1: Lasso, Model 2: Linear, Model 3: Ridge
B) Model 1: Ridge, Model 2: Lasso, Model 3: Linear
C) Model 1: Linear, Model 2: Ridge, Model 3: Lasso
D) Model 1: Lasso, Model 2: Ridge, Model 3: Linear

Answer: A
Explanation:

- **Model 1 (Lasso):** The presence of **exact zeros** in the coefficients indicates L1 regularization, which eliminates irrelevant features.

- **Model 2 (Linear):** No regularization is applied; coefficients are unpenalized and reflect unconstrained least-squares estimates.
- **Model 3 (Ridge):** All coefficients are **small but non-zero**, consistent with L2 regularization shrinking weights toward zero.

Question 3

A healthcare startup is developing a machine learning model to predict patient recovery times based on various medical metrics like age, blood pressure, cholesterol levels, and BMI. The team is debating between using Lasso or Ridge regression for regularization and wants to understand their differences.

Which of the following statements about Lasso and Ridge regression is correct?

Options:
A) Both Lasso and Ridge regression use an L2 penalty term in their cost function.
B) Neither uses penalty terms in their cost function. C) Both Lasso and Ridge regression use an L1 penalty term in their cost function.
D) Lasso regression uses an L1 penalty term, while Ridge uses an L2 penalty term.

Answer: D
Explanation: Lasso regression uses an L1 penalty term (sum of absolute values of coefficients), while Ridge regression uses an L2 penalty term (sum of squared values of coefficients). These penalties control overfitting and improve generalization.

Question 4 (MCQ)

A financial analyst is building a model to predict stock prices based on various market indicators. The dataset has many features, and the analyst wants to ensure that the model doesn't overfit by penalizing large coefficients. They use Ridge regression and experiment with different values for the regularization parameter alpha.

What will happen if they increase the value of alpha in their Ridge regression model?

Options:
A) It will increase the magnitude of the coefficients.
B) It will decrease the magnitude of the coefficients.
C) It will have no effect on the coefficients.
D) It will set some coefficients to exactly zero.

Answer: B
Explanation: Increasing alpha in Ridge regression increases the regularization strength, which penalizes large coefficient values and reduces their magnitude.

Question 5 (MCQ)

An e-commerce company is building a predictive model to estimate delivery times based on factors like distance, traffic conditions, weather, and time of day. The team wants to use a regularization technique that can automatically eliminate irrelevant features while training the model.

Which regularization technique should they choose?

Options:
A) Ridge Regression
B) Lasso Regression
C) Elastic Net Regression
D) Simple Linear Regression

Answer: B
Explanation: Lasso regression uses an L1 penalty term that shrinks some coefficients to exactly zero, effectively removing irrelevant features from the model.

Question 6 (MCQ)

A real estate company is analyzing property prices using a dataset with multiple features such as location, square footage, number of bedrooms, and age of the property. The team notices that some features are highly correlated, such as square footage and number of bedrooms. They want to address multicollinearity while building their predictive model.

Which regularization technique should they use?

Options:
A) Lasso Regression
B) Elastic Net Regression C) Ridge Regression
D) Simple Linear Regression

Answer: C

Explanation:

What is Multicollinearity?

Multicollinearity occurs when two or more features in a dataset are **highly correlated**. For example, in real estate data, "square footage" and "number of bedrooms" often correlate because larger homes tend to have more bedrooms. This can cause instability in regression models, such as:

- Unreliable coefficient estimates (large standard errors).
- Difficulty interpreting feature importance.

How Ridge Regression Addresses Multicollinearity

Ridge regression adds an **L2 penalty term** (sum of squared coefficients) to the loss function:

Loss=Sum of Squared Errors+α×(Sum of Squared Coefficients)

This penalty:

- **Shrinks coefficients** toward zero but **does not eliminate any features**.
- **Distributes the effect** of correlated features across all of them rather than isolating one.

Example:
For correlated features like "square footage" and "number of bedrooms," Ridge regression will:

- Assign **smaller, balanced weights** to both features.
- Stabilize predictions by preventing large swings in coefficients due to data noise.

Why Other Options Are Less Effective

A) Lasso Regression (L1 Penalty)

- **Drawback:** Lasso tends to **select one feature** from a correlated group and discard others (sets their coefficients to zero).
- **Issue:** This eliminates useful information from correlated features, which is undesirable when both contribute meaningfully to the target variable.

B) Elastic Net (L1 + L2 Penalty)

- While Elastic Net combines L1 and L2 penalties, it still prioritizes feature selection (like Lasso).
- **Less optimal** than Ridge when the goal is to retain all correlated features.

C) Simple Linear Regression (No Regularization)

- **Fails to handle multicollinearity:** Coefficients become unstable and sensitive to minor data changes, leading to **overfitting** and unreliable interpretations.

Question 7 (MCQ)

A logistic regression model is used to classify emails as spam or not spam. The model outputs a probability of 0.6 for a particular email.

Question: What is the decision rule in this case?

Options:

A) Predict 0

B) Predict 1

C) Predict 0.5

D) No decision

Correct Answer: B) Predict 1

Explanation:

- Logistic regression **outputs a probability** between **0 and 1**.
- The **default decision threshold** is **0.5**:
- If **probability ≥ 0.5**, predict **1** (spam).
- If **probability < 0.5**, predict **0** (not spam).

Question 8

A logistic regression model computes a **linear combination (z)** of input features and applies the **sigmoid function** to convert **z into a probability**. For a particular input, **z = 1**.

Question:

What is the probability using the sigmoid function? **(Round to 2 decimal places)**

Correct Answer: 0.73

To find the probability, substitute **z = 1** into the formula.

Explanation: The **sigmoid function** is defined as: 1/ {1 + e^{-z}}

For z = 1, 1/ {1 + e^{-1}}

Thus, the **probability is 0.73** when **z = 1**.

Question 9

A company uses multivariate regression to predict salaries: Salary = 25,000 + 3,000 × (Years of Experience) + 6,000 × (Education Level)

where (Education Level: 1 = Bachelor's, 2 = Master's, 3 = PhD)

What is the predicted salary for an employee with 4 years of experience and a Master's degree?

A) $47,000

B) $55,000

C) $50,000

D) $49,000

Answer: D) $49,000

Explanation:

Salary = 25,000 + 3,000 × 4 + 6,000 × 2

= 25,000 + 12,000 + 12,000

= **$49,000**.

Question 10

In polynomial regression, what role does the degree of the polynomial play?**

A) It determines the complexity of the model

B) It is always set to 1

C) It represents the number of independent variables

D) It is always the same as the number of observations

Answer: A

Explanation: Higher-degree polynomials allow the model to fit more complex relationships.

Question 11

What happens when we apply Linear Regression to non-linear data?

A) The model performs well

B) The model fails to capture patterns, leading to high errors

C) The model learns non-linear pattern

D) None of the above

Answer: B) The model fails to capture patterns, leading to high errors

Explanation: Linear Regression assumes a linear relationship between variables, so it fails to model non-linear patterns.

Question 12

Compute MSE for a model which made some predictions:

Actual Values: [10, 20, 30, 40]

Predicted Values: [12, 16, 32, 36]

A) 4

B) 6

C) 8

D) 10

Answer: D) 10

Explanation: The MSE is calculated as:

MSE = {(10-12)^2 + (20-16)^2 + (30-32)^2 + (40-36)^2}/{4}

MSE = {4 + 16 + 4 + 16}/{4} = 10

Question 13

An economist is modeling the relationship between advertising spend and revenue generation for a company using a linear regression approach. **Which equation represents this model? Options:**
A) y=mx+b B) y=ax^2+bx+cy C) y=1/(1 + e^(-z)) D) y=e^x

Answer: A
Explanation: Linear regression models are represented by y=mx+b, where m is the slope and b is the y-intercept.

Question 14

A machine learning researcher is training a polynomial regression model using scikit-learn. The researcher provides input features (X) and target values (Y) to train the model using the following function:

What does the model.fit(X, Y) function do?

Options:
A) Predicts the output for given inputs.
B) Plots a graph of input vs output.
C) Trains the model using input and output data.
D) Calculates statistical metrics for evaluation.

Answer: C
Explanation: The model.fit(X, Y) function trains the model using the provided input (X) and output (Y).

Question 15

A logistic regression model is used to predict whether a customer will purchase a product based on their browsing history. The model outputs a probability score using the sigmoid function.

Question: What is the primary purpose of the sigmoid function in this model?

Options:

A) To normalize data

B) To convert scores into probabilities

C) To minimize errors

D) To calculate weights

Correct Answer: B) To convert scores into probabilities

Explanation:

Logistic regression computes a **linear combination of input features** (i.e., z = wX + b).
The **sigmoid function** is then applied to **map this score to a probability between 0 and 1**:

sigmoid function = $1 / \{1 + e^{-z}\}$

This **probability** represents the likelihood of the positive class (e.g., customer purchasing a product).