Algorithm 1: Linear Regression

Input: Matrix of predictor variables X (with N observations and M predictors); Response variable Y (with N observations)

Output: Estimated parameters $\hat{\boldsymbol{\theta}}$ (including the intercept)

1 Add a column of ones to X to account for the intercept term (augmented matrix):

$$X_{\text{a}} = [\mathbf{1} \,|\, X]$$

2 Calculate the vector of parameters $\hat{\boldsymbol{\theta}}$ using the normal equation:

$$\hat{\boldsymbol{\theta}} = (X_{\text{a}}^T X_{\text{a}})^{-1} X_{\text{a}}^T Y$$

3 **return** $\hat{\boldsymbol{\theta}}$, which includes both the intercept θ_0 and coefficients $\theta_1, \dots, \theta_M$

Algorithm 2: k-Fold Cross Validation for Linear Regression

Input: Predictor variables X (matrix of size $N \times M$); Response variable y (vector of size $N \times 1$); Number of folds k

Output: Average Mean Squared Error (MSE) across all folds

Divide the data into k equal-sized folds

Initialize the total mean squared error to zero

for *each fold i from* 1 *to* k **do**

  Use the i-th fold as the test set and the remaining data as the training set

  Form the training matrix X_{train} by including all predictor variables from the training set

  Add a column of ones to X_{train} to account for the intercept term

  Compute the model parameters $\hat{\theta}_i$ using the normal equation:

$$\hat{\theta}_i \leftarrow (X_{train}^T X_{train})^{-1} X_{train}^T y_{train}$$

  Form the test matrix X_{test} by including all predictor variables from the test set

  Add a column of ones to X_{test} to account for the intercept term

  Predict the response variable for the test set:

$$\hat{y}_{test} \leftarrow X_{test} \hat{\theta}_i$$

  Calculate the errors in the test set:

$$errors \leftarrow y_{test} - \hat{y}_{test}$$

  Compute the mean squared error (MSE) for the test set:

$$\text{MSE}_i \leftarrow \frac{1}{|test|} \sum errors^2$$

  Add the MSE of the current fold to the total mean squared error

Calculate the average MSE by dividing the total MSE by the number of folds:

$$\text{CV}_{(k)} \leftarrow \frac{1}{k} \sum_{i=1}^{k} \text{MSE}_i$$

return the average MSE

Algorithm 3: Bootstrap Resampling for Linear Regression

Input: Predictor variables X (matrix of size $N \times M$); Response variable y (vector of size $N \times 1$); Number of bootstrap samples B

Output: Distributions of regression coefficient estimates and intercept estimates

Initialize lists to store coefficient and intercept estimates

for *each bootstrap sample* b *from* 1 *to* B **do**

  Generate a sample by randomly sampling with replacement from the original data

  Let X_{bs} and y_{bs} be the bootstrap sample of predictor variables and response variable respectively

  Compute the regression coefficients $\hat{\theta}_{bs}$ for the bootstrap sample using:

$$\hat{\theta}_{bs} = (X_{bs}^T X_{bs})^{-1} X_{bs}^T y_{bs}$$

  Compute the intercept $\hat{\theta}_{0,bs}$ as:

$$\hat{\theta}_{0,bs} = \bar{y}_{bs} - X_{bs}^T \hat{\theta}_{bs}$$

  Append the regression coefficients and intercept estimates to their respective lists

Calculate the mean and standard deviation of the regression coefficients and intercept estimates

return the distributions of the regression coefficients and intercept estimates