Report 2

Name	ID
Omar Eldesoukey	202202155
Abdallah Saed	202201933
Alhassan	202200681
Abdulrahman Omar	202202254

Theoretical definition:

Ridge Regression:

This model solves a regression model where the loss function is the linear least square's function and regularization is given by the l_2-norm. Also known as Ridge Regression or Tikhonov regularization. This estimator has built-in support for multi-variate regression (i.e., when y is a 2d-array of shape (n_samples, n_targets)).

Linear least squares with l2 regularization.

Minimizes the objective function:

$$||y - Xw||_2^2 + alpha * ||w||_2^2$$

$$\text{Objective Function (Ridge)} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^{p} \beta_j^2$$

$$\textbf{Objective function} \quad J(\underline{W}) = \frac{1}{2}[\ (\underline{Y} - \hat{\underline{Y}})^T (\underline{Y} - \hat{\underline{Y}}) + \lambda\ W^T W\ ]$$

Regularizes are based on: Penalize large values of W (model parameters) aiming better generalization of the model.

The Regularization term is added to the objective loss function with a customized weigh factor (do regularization to what extent).

Penalization may be applied on the **Squared Norm** of the Weight Vector

This is called L_2 regularization – RIDGE Regression

1. **Linear vs Ridge Regression**:
   - Linear Regression minimizes the residual sum of squares without regularization.
   - Ridge Regression adds an L2 regularization term (alpha) that shrinks coefficients, which helps prevent overfitting.
2. **Alpha in Ridge Regression**:
   - Higher alpha values lead to stronger regularization (simpler models), while lower values let the model fit the data more closely (but risk overfitting).

Bros of Ridge Regression:

- It protects the model from overfitting.

- It does not need unbiased estimators.
- There is only enough bias to make the estimates reasonably reliable approximations to the true population values.
- It performs well when there is a large multivariate data with the number of predictors (p) larger than the number of observations (n).
- The ridge estimator is very effective when it comes to improving the least-squares estimate in situations where there is multicollinearity.
- Model complexity is reduced.

Cons of Ridge Regression:

- It includes all the predictors in the final model.
- It is not capable of performing feature selection.
- It shrinks coefficients towards zero.
- It trades variance for bias.

LASSO Regression:

Lasso regression is a regularization technique that applies a penalty to prevent overfitting and enhance the accuracy of statistical models

Lasso regression also known as L1 regularization—is a form of regularization for linear regression models. Regularization is a statistical method to reduce errors caused by overfitting on training data

$$\text{Objective Function (Lasso)} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^{p} |\beta_j|$$

Pros of LASSO Regression

1. **Feature Selection:** Lasso effectively performs both shrinkage and variable selection, which is vital when sifting through numerous predictors. By penalizing the absolute size of coefficients, it can reduce irrelevant variables to zero, enhancing model interpretability and improving robustness—essential attributes in a volatility-focused market.
2. **Simplicity and Efficiency:** Given its capability to handle multicollinearity and high-dimensional data, lasso can streamline complex models. In quantitative trading, where speed and precision are paramount, a simplified model can lead to quicker decision-making.
3. **Regularization:** The inherent regularization helps minimize overfitting, allowing your strategy to generalize better to unseen data. This is particularly important in seeking consistent alpha generation without the pitfalls of trading noise.

Cons of LASSO Regression

1. **Bias in Coefficients:** While simplifying models, Lasso can introduce bias in coefficient estimates. This can distort the underlying relationships, potentially impairing the alpha you might extract from your trades.
2. **Sensitivity to Lambda:** The choice of the penalty term (lambda) is critical. A poorly chosen lambda can lead to significant underfitting or overfitting. This can result in the erosion of performance metrics, an unforgiving reality when managing a portfolio.
3. **Non-Inclusion of Important Variables:** In cases where highly correlated predictors exist, Lasso may arbitrarily select one and exclude others. This is a risk, especially if a seemingly irrelevant variable holds critical information for predictive accuracy.

Elastic Net Regression

Elastic Net Regression is a regularized regression method that combines the penalties of **Ridge Regression** (L2 penalty) and **Lasso Regression** (L1 penalty) to overcome some of their limitations. It was introduced to provide a solution to the problems faced by Ridge and Lasso when applied individually. Elastic Net is particularly useful when there are multiple features that are highly correlated or when the number of predictors exceeds the number of observations, making it a versatile tool for high-dimensional data.

The Elastic Net combines both **L1** and **L2** penalties, balancing between the **Lasso** and **Ridge** regularization methods. The objective function of Elastic Net is:

$$\text{Objective Function (Elastic Net)} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \lambda_1 \sum_{j=1}^{p} |\beta_j| + \lambda_2 \sum_{j=1}^{p} \beta_j^2$$

Pros of Elastic Net Regression

- **Handling Multicollinearity**: Elastic Net handles multicollinearity better than Lasso by grouping correlated features and selecting the most representative ones.
- **Reducing Model Complexity**: It eliminates irrelevant features more effectively than Ridge regression.
- **Better Bias-Variance Trade-off**: Elastic Net achieves a better trade-off between bias and variance by tuning the regularization parameters, compared to Lasso and Ridge.
- **Flexibility in Application**: It can be applied to various types of models, such as **linear, logistic, or Cox regression**

Cons of Elastic Net Regression

- **Higher Computational Cost**: Elastic Net requires more computational resources and time due to the need to tune two regularization parameters and perform cross-validation.
- **Performance Issues with Uncorrelated Features**: It may not perform optimally when there is no correlation between features or when the number of features is much smaller than the number of observations, potentially losing predictive power or introducing bias.
- **Interpretability Challenges**: Elastic Net may select a large number of features with small coefficients or a small number of features with large coefficients, making the model less interpretable.

References:

https://scikit-learn.org/1.5/modules/generated/sklearn.linear_model.Ridge.html

https://www.tutorialspoint.com/scikit_learn/scikit_learn_ridge_regression.htm
https://docs.esmvaltool.org/en/v2.9.0/api/esmvaltool.diag_scripts.mlr/models.ridge.html

https://online.stat.psu.edu/stat857/node/155/

https://www.ibm.com/topics/lasso-regression

https://www.geeksforgeeks.org/what-is-lasso-regression/

https://www.ibm.com/docs/en/spss-statistics/saas?topic=features-linear-elastic-net-regression

https://scikit-learn.org/1.5/modules/generated/sklearn.linear_model.ElasticNet.html

https://academic.oup.com/jrsssb/article/67/2/301/7109482