

## PENERAPAN REGULARISASI LASSO DAN RIDGE DALAM MENGATASI MASALAH MULTIKOLINEARITAS PADA MATRIKS KOVARIANS

Muhammad Rafli Lubis<sup>1</sup>, Sutarman<sup>2</sup>

Email: [raflimuhammad080103@gmail.com](mailto:raflimuhammad080103@gmail.com)<sup>1</sup>, [sutarman@usu.ac.id](mailto:sutarman@usu.ac.id)<sup>2</sup>

Universitas Sumatera Utara

### ABSTRAK

Matriks kovarian merupakan hubungan ukuran seberapa jauh dua atau lebih variabel acak bervariasi bersama. dalam konteks pemodelan dan pemrosesan data yang lebih kompleks matriks kovarians dapat mengalami multikolinearitas yang sangat tinggi dan memiliki nilai determinan yang amat sangat kecil atau bahkan nol sehingga menjadi singular. Penelitian ini bertujuan mengatasi permasalahan singularitas pada matriks kovarians yang disebabkan oleh adanya multikolinearitas dengan menerapkan penalti regularisasi L1 (LASSO) yang akan dibandingkan dengan metode regularisasi L2 (Ridge). Regularisasi ridge akan menambahkan penalti yang kemudian diaplikasikan terhadap nilai diagonal matriks kovarians sehingga dapat mengurangi sifat ketergantungan linear (Linear Dependencies) dan akan meningkatkan nilai determinan dari matriks kovarians itu sendiri. Sedangkan dalam LASSO akan menambahkan nilai penalti terhadap nilai absolut dari semua koefisien. Data yang digunakan merupakan matriks kovarians yang di bangkitkan dengan dimensi dan struktur yang bervariasi. metode LASSO mampu memberikan determinan yang lebih baik namun tertinggal cukup jauh jika dibandingkan dengan metode regularisasi Ridge berkisar  $1e^{(+1)}$ - $1e^{(+5)}$ , dan metode LASSO tidak dapat secara langsung memaksimalkan nilai rank matriks sangat berbeda jauh dibandingkan metode ridge yang mampu memaksimalkan nilai rank matriks kovarians yang singular. Dari hasil tersebut dapat dilihat bahwa metode ridge terbukti lebih baik dibandingkan LASSO dalam meningkatkan determinan dan nilai rank matriks dari matriks kovarians yang singular.

**Kata Kunci:** LASSO, Multikolinearitas, Regularisasi, Ridge, Masalah Singular Dalam Kovarians Matriks.

### ABSTRACT

*The covariance matrix is a relationship measure of how far two or more random variables vary together. In the context of modelling and processing more complex data, the covariance matrix can experience very high multicollinearity and have a determinant value that is very small or even zero so that it becomes singular. This research aims to overcome the problem of singularity in the covariance matrix caused by multicollinearity by applying the L1 (LASSO) regularization penalty which will be compared with the L2 (Ridge) regularization method. Ridge regularization will add a penalty which is then applied to the diagonal value of the covariance matrix so as to reduce the dependency properties and will increase the determinant value of the covariance matrix itself. Meanwhile, LASSO will add a penalty value to the absolute value of all coefficients. The data used is a covariance matrix that is generated with varying dimensions and dependency structures. LASSO method is able to provide a better determinant but lags far enough when compared to the ridge method around  $1e^{(+1)}$ - $1e^{(+5)}$  and the LASSO method cannot directly maximize the rank value of the matrix. It is very different from the ridge method which is able to maximize the rank value of the covariance matrix that is singular. From these results it can be seen that the ridge method is proven to be better than LASSO in increasing the determinant and rank matrix values of a singular covariance matrix.*

**Keywords:** LASSO, Multicollinearity, Regularization, Ridge, Singular Problem In Covariance Matrix.

## **PENDAHULUAN**

Regularisasi merupakan teknik yang diterapkan dalam pemodelan statistik dan pembelajaran mesin untuk mencegah model mempelajari data pelatihan dengan terlalu baik (overfitting) serta meningkatkan kemampuan generalisasi model (Bondarenko, 2023). Regularisasi menerapkan penalti tambahan pada model yang bertujuan untuk mengontrol kompleksitas model dan menghindari ketergantungan berlebihan pada fitur atau pola tertentu dalam data pelatihan sehingga tercapai keseimbangan antara menyesuaikan data pelatihan dan menggunakannya dengan baik ke data baru (Ayyildiz et al., 2012).

Regularisasi L1 (LASSO) dan L2 (Ridge) merupakan teknik regularisasi yang paling umum digunakan. L1 menambahkan jumlah nilai absolut koefisien model ke fungsi kerugian, mendorong ketersebaran dan pemilihan fitur. Regularisasi L2 menambahkan jumlah nilai kuadrat dari koefisien model, yang memungkinkan koefisien lebih kecil namun bukan nol. Penggunaan regularisasi L1 (LASSO) dan L2 (Ridge) tidak hanya membantu dalam mengontrol kompleksitas model, tetapi juga memberikan wawasan yang berharga tentang hubungan antar variabel dengan melibatkan matriks kovarians (Hoerl dan Kennard, 1970).

Dalam statistik, matriks kovarian merupakan hubungan ukuran seberapa jauh dua atau lebih variabel acak bervariasi bersama, dalam kasus di mana ada lebih dari dua variabel acak, matriks ini biasanya disusun dalam bentuk matriks simetris. Matriks ini dapat digunakan untuk melihat bagaimana variabel-variabel tersebut berhubungan satu sama lain, apakah mereka cenderung bergerak bersamaan (korelasi positif), bergerak berlawanan (korelasi negatif), atau tidak memiliki hubungan yang signifikan (Oyeyemi dan Ipinoyomi, 2009). Dalam konteks pemodelan dan pemrosesan data yang lebih kompleks, matriks kovarians dapat mengalami ill-condition yaitu non-invertible matrices atau yang lebih sering dikenal dengan matriks singular (Ledoit dan Wolf, 2004).

Matriks singular merupakan metriks yang tidak memiliki invers atau matriks non-invers (singular matrices or non-invertible matrices). Matriks ini memiliki determinan yang sama dengan nol sehingga tidak dapat diubah kembali ke matriks identitas, Ketika suatu matriks tidak memiliki invers, itu berarti sistem persamaan linear yang terkait mungkin tidak memiliki solusi unik atau bahkan tidak memiliki solusi sama sekali. Oleh karena itu, matriks singular dapat menyebabkan ketidakmungkinan untuk menyelesaikan masalah dengan metode invers. Matriks singular biasanya disebabkan oleh multikolinearitas yang tinggi. Keberadaan multikolinearitas akan mempengaruhi nilai estimator sehingga memiliki varian yang besar dan menghasilkan tingkat akurasi dari estimasi akan menurun. (Sukmono dan Subiyanto, 2014).

Dalam menghadapi masalah singularitas pada matriks kovarians, metode Cholesky Decomposition lebih efektif dibandingkan dengan metode Eigenvalue Decomposition dan singular value Decomposition (Ayyildiz et al., 2012). Ketiga metode tersebut menggunakan konsep dekomposisi yang akan mengurai atau mereduksi dimensi dari matriks kovarians, lebih lanjut Bondarenko (2023) menyelesaikan masalah singularitas pada operator diferensial dengan menggunakan regularisasi Mirzoev dan Shkalikov serta mendapatkan kesimpulan bahwa regularisasi memberikan solvability dan stabilitas yang baik dalam menghadapi masalah singularitas pada matriks.

## **METODE PENELITIAN**

Untuk menyelesaikan permasalahan singularitas pada matriks varians-kovarians digunakan metode penalti regularisasi Least Absolute Shrinkage and Selection Operator (LASSO) dan Ridge, Penalti LASSO dan Ridge yang diterapkan terhadap matriks varians - kovarians diharapkan dapat meningkatkan nilai rank matriks dan juga determinan yang bernilai 0 ataupun sangat kecil hingga mendekati 0. Ketika penerapan penalti regularisasi selesai, maka akan dibuat perbandingan hasil dari kedua metode regularisasi dan juga

kesimpulan dari penelitian ini. Adapun langkah-langkah penelitian ini disusun sebagai berikut :

#### 1. Studi Literatur

Langkah pertama yang dilakukan yaitu studi literatur melalui cara mencari literatur seperti teori dari buku, jurnal, artikel penelitian terdahulu yang relevan dengan permasalahan kesingularitasan pada matriks varians – kovarians dan metode regularisasi yang meliputi LASSO dan Ridge. Beberapa literatur yang terkumpul dapat dijadikan sebagai pedoman dalam melakukan penelitian ini.

#### 2. Pembangkitan data

#### 3. Pengolahan data

- a. Pengelompokkan data berdasarkan struktur dependensi dan rank matriks
- b. Menentukan nilai determinan dan rank matriks dari data bangkitan pertama dan kedua
- c. Menerapkan penalti regularisasi LASSO dan Ridge
- d. Membandingkan hasil nilai determinan dan perubahan rank matriks

#### 4. Membuat Kesimpulan dan saran

### **HASIL DAN PEMBAHASAN**

Penerapan metode regularisasi LASSO dan Ridge pada data matriks bangkitan yang mengalami multikolinearitas, yang sering kali mengakibatkan singularitas pada matriks. Data yang digunakan adalah data simulasi yang dibangkitkan dari matriks kovarians simetris dan singular. Dalam pembangkitan ini, kolom independen digabungkan dengan kolom dependen untuk menghasilkan matriks singular, di mana persentase kolom linear dependent meningkat pada setiap skenario, dimulai dari 30%, kemudian 40%, hingga 50%.

Pada skenario pertama, matriks yang dibangkitkan memiliki 30% kolom linear dependent. Untuk skenario kedua, persentase kolom dependent dinaikkan menjadi 40%, dan skenario ketiga memiliki 50% kolom dependent yang digabungkan dengan kolom independen. Proses pembangkitan dan analisis data dilakukan menggunakan bahasa pemrograman R pada platform berbasis cloud, Google Colab. Data lengkap dari matriks bangkitan ini terdapat pada Lampiran 5, yang menyajikan nilai-nilai matriks pada setiap skenario.

Analisis berikutnya melibatkan perhitungan nilai skalar kovarians untuk setiap variabel dalam matriks bangkitan, yang kemudian disusun dalam bentuk matriks varians-kovarians. Matriks ini mengalami masalah multikolinearitas yang serius hingga menyebabkan singularitas, yang dapat dilihat pada determinan matriks yang sangat kecil atau mendekati nol, sebagaimana tercantum pada Tabel 4.3.

Untuk mengatasi masalah singularitas, metode regularisasi seperti LASSO dan Ridge diterapkan. Regularisasi LASSO memberikan penalti terhadap nilai absolut koefisien dalam matriks, sementara Ridge menambahkan penalti pada diagonal elemen matriks. Kedua metode ini bertujuan untuk mengurangi nilai koefisien berlebih yang menyebabkan multikolinearitas, sehingga memperbaiki singularitas dan memastikan matriks kovarians menjadi invertible.

Pada penerapan LASSO, nilai determinan matriks meningkat secara signifikan, tetapi tidak sebaik metode Ridge. LASSO memberikan bias kecil pada matriks varians-kovarians melalui parameter penalti ( $\lambda$ ), yang diatur sebesar 0.1 dalam penelitian ini. Hasil determinan menggunakan metode LASSO untuk setiap skenario tercantum pada Tabel 4.4.

Metode Ridge, di sisi lain, menggunakan pendekatan penalti yang berbeda, di mana penalti diterapkan pada elemen diagonal dari matriks kovarians. Metode ini secara efektif meningkatkan nilai determinan dan rank matriks, sehingga memberikan hasil yang lebih stabil pada matriks singular dengan multikolinearitas tinggi. Determinan hasil Ridge tercantum pada Tabel 4.5 dan menunjukkan peningkatan signifikan dibandingkan LASSO.

Hasil dari kedua metode ini dibandingkan melalui grafik nilai determinan matriks pada masing-masing skenario. Ridge menunjukkan peningkatan nilai determinan yang lebih konsisten dan signifikan pada setiap skenario dibandingkan LASSO. Grafik perbandingan nilai determinan untuk masing-masing skenario disajikan pada Gambar 4.1 hingga 4.3.

Ridge memiliki keunggulan dalam meningkatkan rank matriks hingga ke rank maksimal, sementara LASSO hanya mampu meningkatkan rank matriks secara terbatas. Hal ini terlihat dari Gambar 4.4 dan 4.5, yang menunjukkan bahwa Ridge mampu mengatasi dependencies lebih efektif, mencapai rank maksimal untuk matriks 20x20 dan 30x30, sedangkan LASSO hanya mampu meningkatkan satu atau dua nilai rank matriks.

Secara keseluruhan, penelitian ini menunjukkan bahwa metode Ridge lebih unggul dibandingkan LASSO dalam mengatasi masalah singularitas pada matriks kovarians dengan multikolinearitas.

## **KESIMPULAN**

1. Berdasarkan pada bab hasil dan pembahasan didapati kesimpulan bahwa metode regularisasi ridge terbukti lebih baik dalam skenario 1, skenario 2, dan skenario 3 sebesar  $1e^{(+1)}-1e^{(+5)}$  dalam meningkatkan nilai determinan dibandingkan metode LASSO dalam simulasi matriks kovarians.
2. Kesimpulan dari uji coba peningkatan nilai rank matriks menunjukkan bahwa metode LASSO secara bertahap meningkatkan rank matriks seiring dengan bertambahnya urutan, namun tetap tidak mencapai nilai rank maksimal. disisi lain, metode ridge berhasil mempertahankan rank maksimal pada setiap urutan, menunjukkan bahwa ridge lebih efektif dalam mempertahankan atau meningkatkan rank matriks dibandingkan dengan LASSO dalam kasus ini.
3. Dari hasil penelitian ini menunjukkan bahwa metode ridge memang lebih baik dan efektif dalam meningkatkan dan memaksimalkan nilai determinan dan peningkatan rank matriks dari pada metode LASSO di dalam kasus matriks kovarians dengan multikolinearitas tinggi dan bahkan singular, namun perlu diperhatikan bahwa baik LASSO maupun ridge tidak memberikan peningkatan nilai determinan yang signifikan jika berbicara didalam skala yang lebih besar, namun metode ini sudah lebih dari cukup untuk mengatasi permasalahan singularitas dan membuat matriks simulasi kovarians awal untuk memiliki nilai determinan dan nilai rank matriks yang lebih baik dan menyelesaikan kasus invertible matriks.

## **Saran**

Terkait dengan perolehan hasil peningkatan nilai determinan dalam matriks kovarians yang singular perlu diadakan penelitian lebih lanjut menggunakan metode dan skenario yang berbeda, kemudian juga diharapkan untuk dapat meningkatkan baik dimensi dan multikolinearitas pada data yang akan diteliti pada penelitian selanjutnya dan diharapkan juga bisa mengembangkan penelitian dengan memberikan nilai outlier pada data.

## **DAFTAR PUSTAKA**

- Ayyildiz E, Gazi V, Wit E. (2012). A Short Note on Resolving Singularity Problems in Covariance Matrices. *International Journal of Statistics and Probability*, 1(2). <https://doi.org/10.5539/ijsp.v1n2p113>
- Bondarenko P. (2023). Regularization and Inverse Spectral Problems for Differential Operators with Distribution Coefficients. *Mathematics*, 11(16), 3455. <https://doi.org/10.3390/math11163455>
- Deprez M, dan Robinson C. (2024). Dimensionality reduction. *Machine Learning for Biomedical Applications*, 105–137. <https://doi.org/10.1016/B978-0-12-822904-0.00010-8>
- Friedberg S, Insel A, Spence E. (2014). *Linear algebra*.
- Friedman J, Hastie T, Höfling H, Tibshirani R. (2007). Pathwise coordinate optimization. *The Annals of Applied Statistics*, 1(2). <https://doi.org/10.1214/07-aos131>

- Hasim J. (2022). Metode Regresi Least Absolute Shrinkage and Selection.
- Hoerl E, Kennard W. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems (Vol. 12, Issue 1).
- Horn A, Johnson R. (2012). Matrix Analysis Second Edition.
- Ifadah a. (2011). Analisis Metode Principal Component Analysis dan Regresi Dalam Mengatasi Dampak Multikolinearitas Dalam Analisis Linear Berganda.
- Ledoit O, Wolf M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88(2), 365–411. [https://doi.org/10.1016/S0047-259X\(03\)00096-4](https://doi.org/10.1016/S0047-259X(03)00096-4)
- Li Q, Shao J. (2015). Regularizing LASSO: A Consistent Variable Selection Method. 25(3), 975–992. <https://doi.org/10.5705/ss>
- Moser K. (1996). Linear Algebra and Related Introductory Topics. *Linear Models*, 1–22. <https://doi.org/10.1016/B978-012508465-9/50001-6>
- Ndii M. (2018). Pemodelan Matematika Dinamika Populasi dan Penyebaran Penyakit: Teori, Aplikasi, dan Numerik. <https://www.researchgate.net/publication/330840391>
- Oyeyemi M, dan Ipinyomi A. (2009). A Robust Method of Estimating Covariance Matrix In Multivariate Data Analysis.
- Patil R, Kim S. (2020). Combination of ensembles of regularized regression models with resampling-based LASSO feature selection in high dimensional data. *Mathematics*, 8(1). <https://doi.org/10.3390/math8010110>
- Pham-Gia T, Thanh DN. (2016). Hypergeometric Functions: From One Scalar Variable to Several Matrix Arguments, in *Statistics and Beyond*. *Open Journal of Statistics*, 06(05), 951–994. <https://doi.org/10.4236/ojs.2016.65078>
- Sakinah R. (2020). Analisis Sistem Rekomendasi Data Rating AIRBNB Menggunakan Inisialisasi Non-Negative Double Singular Value Decomposition Pada Metode Non-Negative Matrix Factorization.
- Simon N, Friedman J, Hastie T, Tibshirani R. (2011). Regularization paths for Cox’s proportional hazards model via coordinate descent. *Journal of Statistical Software*, 39(5), 1–13. <https://doi.org/10.18637/jss.v039.i05>
- Sukmono A. (2014). Penggunaan Partial Least Square Regression (PLSR) untuk Mengatasi Multikolinearitas Dalam Estimasi Klorofil Daun Tanaman Padi dengan Citra Hiperspektral.