



Supervised Learning

#5 Meeting

LASSO Regression

Ferdian Bangkit Wijaya, S.Stat., M.Si
NIP. 199005202024061001

Kelemahan Ridge Regression

Pada sesi sebelumnya, kita melihat bahwa Ridge Regression sangat hebat dalam menangani multikolinearitas. Ia menstabilkan model dengan menyusutkan koefisien-koefisien yang "meledak".

Namun, Ridge memiliki satu keterbatasan:

- Ia akan membuat koefisien menjadi sangat kecil, tetapi tidak akan pernah membuatnya menjadi persis nol.

Artinya, semua variabel prediktor, tidak peduli seberapa tidak relevannya, akan tetap ada di dalam model final.

Dalam skenario di mana kita memiliki ratusan atau ribuan prediktor, kita tidak hanya ingin menstabilkan model, tetapi juga menyederhanakannya dengan membuang variabel yang tidak penting.

Lasso: Shrinkage + Seleksi Fitur

Lasso (Least Absolute Shrinkage and Selection Operator) diciptakan untuk melakukan dua pekerjaan sekaligus:

1. Shrinkage: Sama seperti Ridge, Lasso menyusutkan koefisien untuk mengurangi varians dan menstabilkan model.
2. Seleksi Fitur (Feature Selection): Tidak seperti Ridge, Lasso mampu menyusutkan koefisien dari variabel yang tidak penting hingga menjadi tepat nol, yang secara efektif menghilangkannya dari model.

Ini membuat Lasso menjadi alat yang sangat ampuh untuk membangun model yang lebih sederhana (parsimonious) dan mudah diinterpretasikan.

Fungsi Biaya : RSS + Penalti L1

$$\text{Minimalkan} \left\{ \underbrace{\sum_{i=1}^n (y_i - \hat{y}_i)^2}_{\text{RSS (Kecocokan Model)}} + \underbrace{\lambda \sum_{j=1}^k |\beta_j|}_{\text{Penalti L1 (Kompleksitas Model)}} \right\}$$

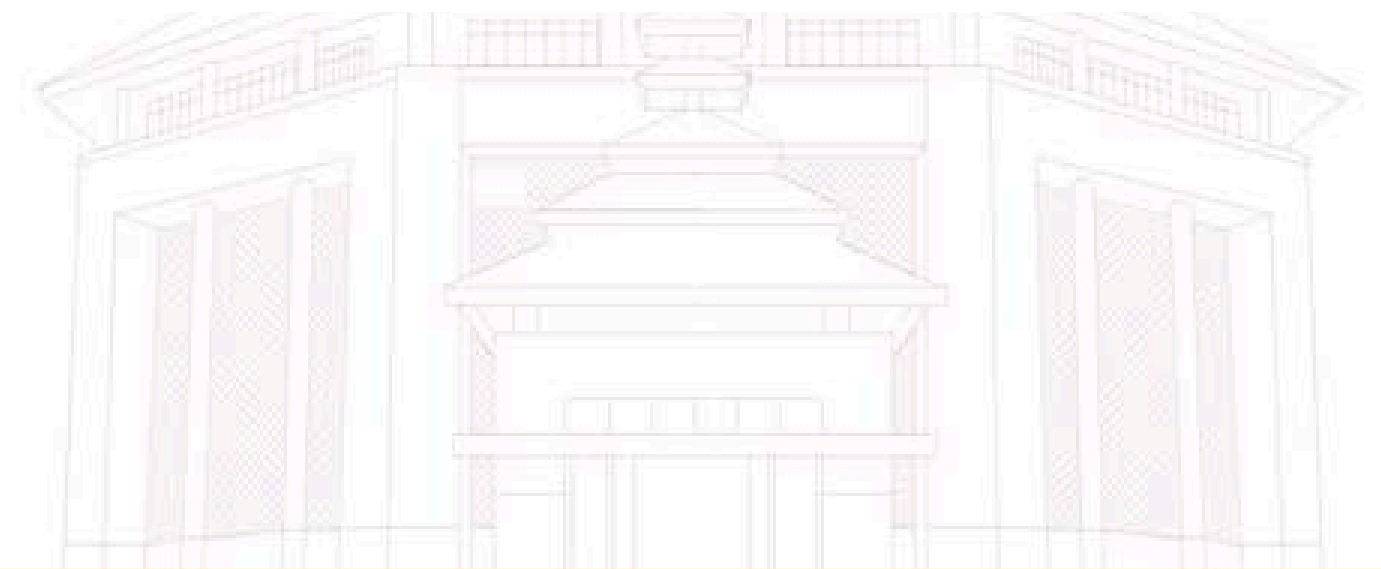
Perhatikan perbedaan kuncinya: $|\beta_j|$ (nilai absolut) menggantikan β_j^2 (nilai kuadrat) dari Ridge.

Penalti L1: $\lambda \sum |\beta_j|$

- $\sum |\beta_j|$ (Jumlah Nilai Absolut Koefisien):
 - Ini adalah "Norma L1" dari vektor koefisien.
 - Hukumannya meningkat secara linear dengan besaran koefisien. Berbeda dengan L2 (Ridge) yang hukumannya meningkat secara eksponensial.
 - Sifat linear inilah yang secara matematis memungkinkan penalti untuk "mengatasi" kekuatan dari RSS dan mendorong koefisien hingga ke titik nol.
- λ (Lambda):
 - Sama seperti di Ridge, λ adalah kenop yang mengontrol kekuatan penalti. Semakin besar λ , semakin banyak koefisien yang akan menjadi nol.

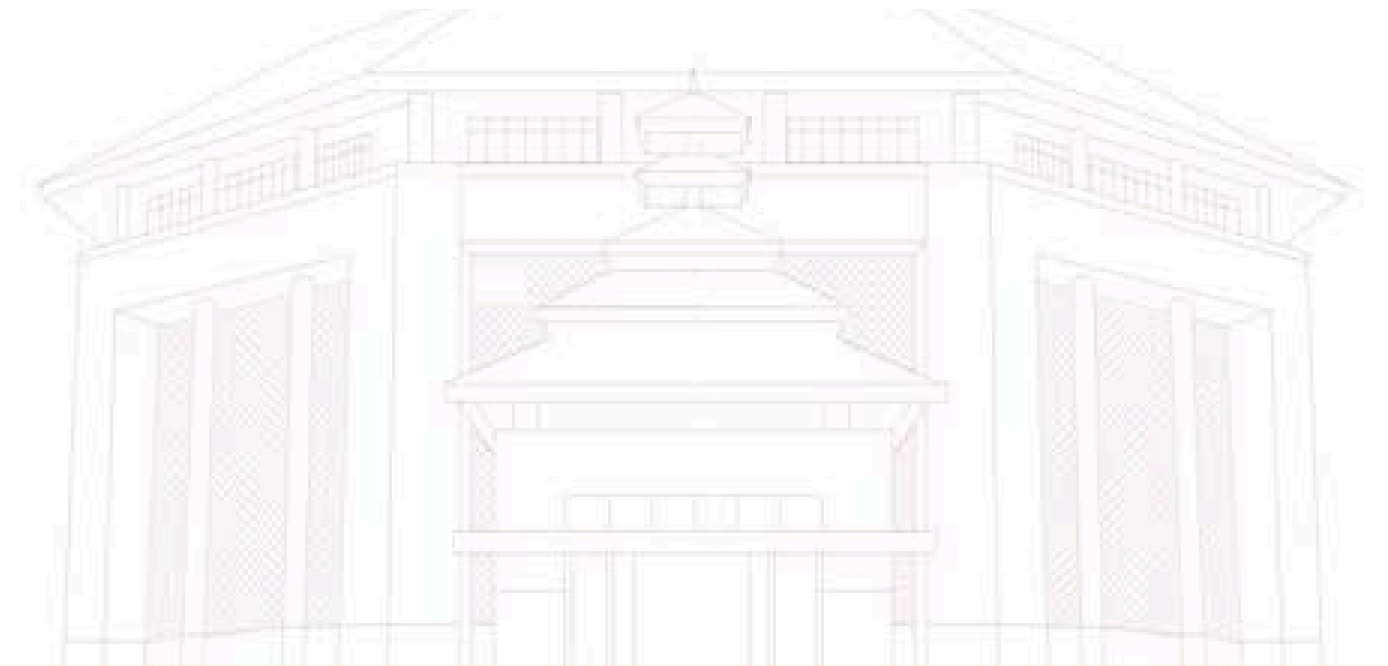
Seleksi Fitur Otomatis

- Kemampuan Lasso untuk membuat koefisien menjadi nol adalah kekuatannya.
- Jika sebuah variabel x_j mendapatkan koefisien $\hat{\beta}_j = 0$, itu berarti model final sama sekali tidak menggunakan variabel tersebut untuk membuat prediksi.
- Lasso secara otomatis "memilih" subset dari prediktor yang paling penting dan "membuang" sisanya.
- Contoh :
- Ini sangat berguna ketika kita memiliki puluhan ribu gen untuk memprediksi suatu penyakit, dan kita yakin hanya beberapa gen saja yang benar-benar relevan.



Penurunan Rumus Pendugaan

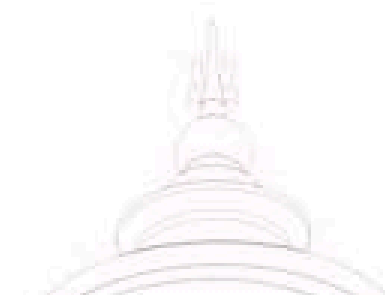
- Berbeda dengan OLS dan Ridge, tidak ada solusi analitis (bentuk-tertutup) untuk Lasso.
- Alasan: Fungsi nilai absolut $f(\beta) = |\beta|$ tidak dapat diturunkan (not differentiable) pada titik $\beta = 0$. Ia memiliki "sudut" yang tajam.
- Konsekuensi: Kita tidak bisa menggunakan kalkulus standar (mengatur turunan = 0 dan selesaikan β) seperti yang kita lakukan untuk Ridge.
- Solusi: Koefisien Lasso harus ditemukan menggunakan algoritma optimisasi numerik yang bersifat iteratif, seperti **Coordinate Descent**.



Peran λ

- Di Lasso, λ tidak hanya mengontrol besarnya penyusutan, tetapi juga jumlah fitur yang dipilih.
- Jika $\lambda = 0$: Lasso menjadi identik dengan OLS.
- Jika λ meningkat:
 - Semakin banyak koefisien yang akan "didorong" menjadi nol.
 - Model menjadi semakin sederhana.
- Tujuan: Sama seperti Ridge, kita menggunakan Validasi Silang (Cross-Validation) untuk menemukan nilai λ yang memberikan keseimbangan terbaik antara akurasi dan kesederhanaan.

Case Method : LASSO

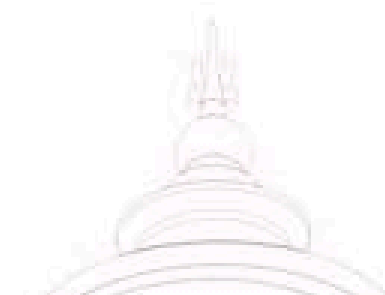


Download Raw Data LASSO pada link di bawah ini :

LINK

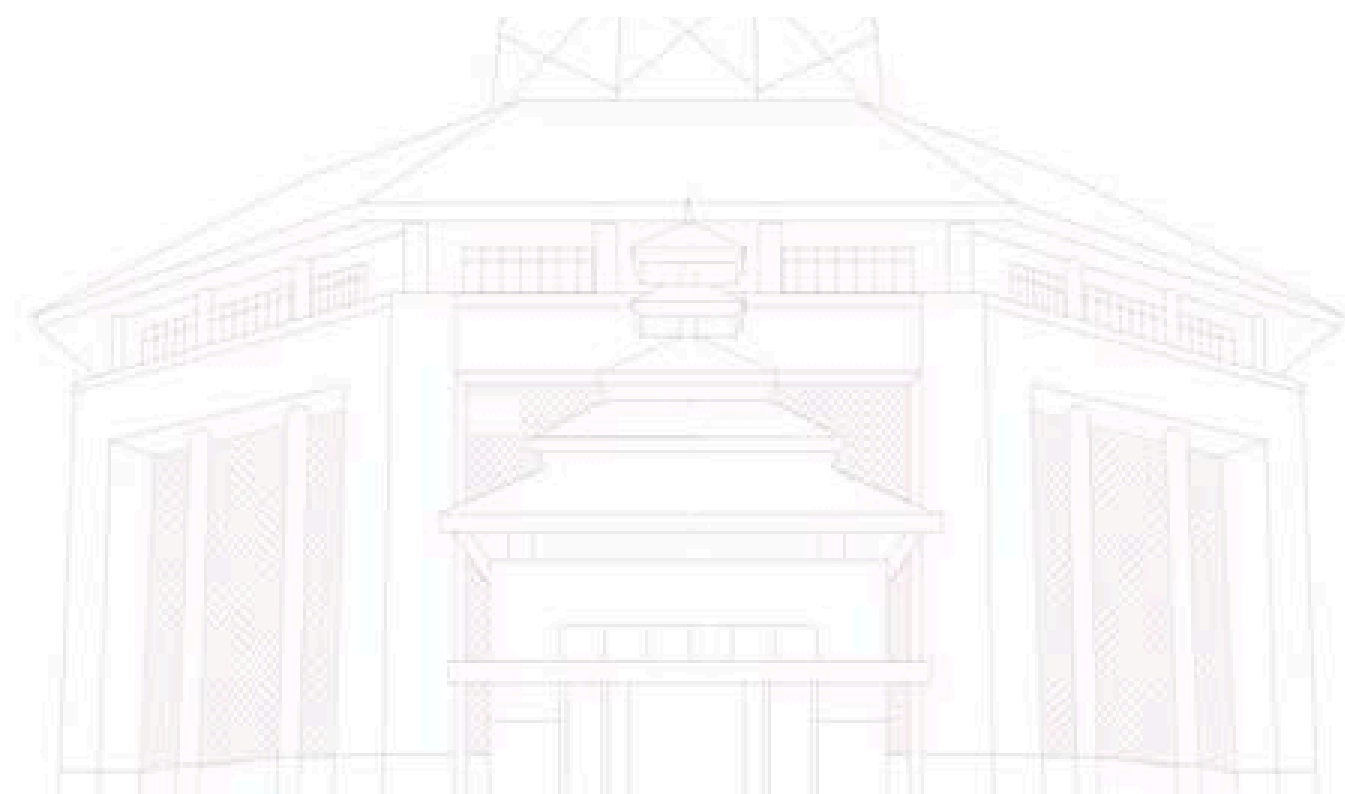


Case Method : LASSO in R

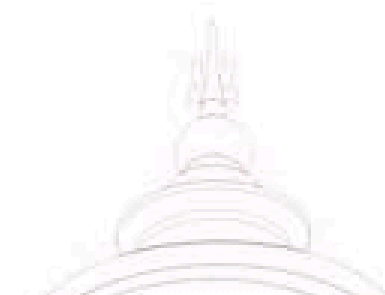


Running Data LASSO pada link di bawah ini :

LINK



Case Method : LASSO in Python



Running Data LASSO pada link di bawah ini :

LINK





SEE YOU NEXT WEEK !

Ferdian Bangkit Wijaya, S.Stat., M.Si

NIP. 199005202024061001

ferdian.bangkit@untirta.ac.id