UNIVERSITAS SULTAN AGENG TIRTAYASA

UNTIRTA Jawara

Supervised Learning

#4 Meeting

Ridge Regression

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

OLS (Ordinary Least Square)

UNIVERSITAS SULTAN AGENG TIRTAYASA

UNTIRTA Jawara

- Ordinary Least Squares (OLS) bertujuan meminimalkan RSS. Namun, pendekatan ini menjadi tidak stabil saat menghadapi:

1. Multikolinieritas:

- Variabel prediktor saling berkorelasi tinggi.
- OLS "bingung" menentukan kontribusi unik setiap variabel.

2. Overfitting:

- Model terlalu "menghafal" data training, termasuk noise-nya.
- Terjadi saat model terlalu kompleks (misalnya, jumlah prediktor sangat banyak).

Kedua masalah ini menyebabkan estimasi koefisien (β) yang tidak stabil dan memiliki varians yang sangat tinggi.

Multikolinieritas pada OLS

- Secara matematis, multikolinieritas menyebabkan matriks (X^TX) menjadi hampir singular (determinannya mendekati nol).

Konsekuensi:

- Invers matriks, $(X^TX)^{-1}$, akan memiliki elemen-elemen dengan nilai yang sangat besar.
- Ini menyebabkan koefisien $\hat{\beta}$ yang dihitung menjadi sangat besar dan tidak stabil.
- Tanda koefisien (+/-) bisa berlawanan dengan teori dan sangat sensitif terhadap perubahan kecil pada data.
- Model menjadi tidak dapat diandalkan baik untuk interpretasi maupun prediksi.

Filosofi Ridge Regression

- Ridge Regression memperkenalkan sebuah trade-off (pertukaran).
- Daripada mencari model dengan RSS terendah mutlak yang berisiko tidak stabil, lebih baik mencari model yang sedikit bias (RSS sedikit lebih tinggi) tetapi jauh lebih stabil (varians koefisiennya rendah).
- Ridge mencapai ini dengan memberikan "batasan" atau "hukuman" pada koefisien, mencegahnya tumbuh terlalu besar.

UNIVERSITAS SULTAN AGENG TIRTAYASA

Standarisasi Data

UNTIRTA Jawara

Secara praktik, sangat sangat disarankan (bahkan dianggap wajib) untuk melakukan standarisasi data sebelum menerapkan Ridge (dan juga Lasso). Sebaiknya tidak menggunakan data mentah secara langsung jika variabel-variabel memiliki skala yang berbeda.

Sistem "denda" Ridge tidak peduli dengan skala asli variabel. Ia hanya melihat besaran angka koefisien β. Ia akan memberikan denda yang jauh lebih besar pada β_2 yang angkanya besar, dan denda yang sangat kecil pada β_1 yang angkanya kecil.

Ini tidak adil. Variabel x_1 seolah-olah mendapat "perlakuan istimewa" hanya karena unit pengukurannya (meter) menghasilkan koefisien yang kecil.

Solusinya: Dengan standarisasi, kita mengubah semua variabel prediktor ke dalam skala yang sama (biasanya dengan rata-rata 0 dan standar deviasi 1). Ini seperti mengubah semua variabel menjadi "skor-z". Setelah distandarisasi, besaran koefisien β yang dihasilkan akan benar-benar mencerminkan kekuatan prediktif relatif dari setiap variabel, bukan lagi skala unit pengukurannya.

Dengan begitu, penalti Ridge dapat diterapkan secara adil dan efektif.

Fungsi Biaya Baru

Ridge memodifikasi fungsi biaya OLS dengan menambahkan suku penalti L2.

$$\text{Minimalkan}\{ \underbrace{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}_{\text{RSS (Kecocokan Model)}} + \underbrace{\lambda\sum_{j=1}^{k}(\beta_j)^2}_{\text{Penalti L2 (Kompleksitas Model)}} \}$$

Model sekarang harus menyeimbangkan dua hal:

1. Menjaga agar tetap cocok dengan data (RSS rendah).
2. Menjaga agar total kuadrat koefisiennya tetap kecil (Penalti rendah).

Pinalti L2

- Fokus pada Penalti L2: $\lambda \sum \beta_j^2$

- $\sum(\beta_j)^2$ (Jumlah Kuadrat Koefisien):
- Ini adalah "Norma L2" dari vektor koefisien (tidak termasuk intersep β_o).
- Sifat kuadrat ini memberikan hukuman yang jauh lebih besar pada koefisien yang bernilai besar (misal, β=4 dihukum 16 kali, β=2 dihukum 4 kali).

- λ (Lambda):
- Disebut Tuning Parameter atau parameter regularisasi.
- Ini adalah "kenop" yang kita atur untuk mengontrol seberapa kuat penalti yang diberikan. Semakin besar λ, semakin besar penyusutan (shrinkage) koefisien.

Penduga Parameter

- Kita dapat menurunkan solusi untuk β Ridge secara analitis.

- Fungsi Biaya dalam Bentuk Matriks:
- $J(\beta) = (y - X\beta)^T(y - X\beta) + \lambda\beta^T\beta$

$$\text{Minimalkan}\{\ \underbrace{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}_{\text{RSS (Kecocokan Model)}} + \underbrace{\lambda\sum_{j=1}^{k}(\beta_j)^2}_{\text{Penalti L2 (Kompleksitas Model)}}\ \}$$

- Ambil Turunan (Gradient) terhadap β:
- $\partial J/\partial\beta = -2X^Ty + 2X^TX\beta + 2\lambda\beta$

- Atur Turunan = 0 dan Selesaikan untuk β:
- $-2X^Ty + 2X^TX\beta + 2\lambda\beta = 0(X^TX + \lambda I)\beta = X^Ty$

- Solusi Akhir (Penduga Ridge):
- $\hat{\beta}$_ridge $= (X^TX + \lambda I)^{-1} X^Ty$

Penduga Parameter

- $\hat{\beta}$_ridge = $(X^TX + \lambda I)^{-1} X^Ty$
- Bagaimana $+\lambda I$ Bekerja?

- Perbandingan formula menunjukkan solusi elegannya:
- OLS: $\hat{\beta}$_ols = $(X^TX)^{-1} X^Ty$
- Ridge: $\hat{\beta}$_ridge = $(X^TX + \lambda I)^{-1} X^Ty$
- Penambahan λI (sebuah konstanta positif λ pada diagonal matriks X^TX) secara matematis:
- Menjamin matriks $(X^TX + \lambda I)$ dapat di-invers, bahkan jika X^TX hampir singular.
- Menstabilkan proses inversi, mencegah angka-angka di dalamnya "meledak".
- Menghasilkan koefisien $\hat{\beta}$ yang menyusut (shrunken) dan lebih stabil.

Lambda

- Peran Krusial λ (Lambda)

- Nilai λ mengontrol seberapa besar penyusutan yang terjadi.
- Jika λ = 0:
- Penalti hilang.
- $\hat{\beta}$_ridge menjadi identik dengan $\hat{\beta}$_ols.

- Jika λ meningkat:
- Penalti semakin kuat.
- Koefisien $\hat{\beta}$ akan semakin menyusut mendekati nol.

- Jika λ sangat besar (→ ∞):
- Semua koefisien akan menjadi mendekati nol, menghasilkan model yang sangat sederhana (underfitting).

Optimasi Lambda

- Bagaimana Menemukan λ Terbaik? Validasi Silang

- Kita tidak memilih λ secara acak. Metode standar untuk menemukan λ optimal adalah Validasi Silang (Cross-Validation).

- Bagi Data: Data training dibagi menjadi K bagian (misal, 10-Fold CV).

- Iterasi:
- Model dilatih pada K-1 bagian data dengan berbagai nilai λ.
- Model diuji pada 1 bagian data yang disisihkan (validation set).
- Pilih λ Terbaik: Ulangi proses ini K kali. Pilih nilai λ yang secara rata-rata menghasilkan error prediksi terendah (misal, MSE terendah) pada validation set.

Case Method : Ridge

Download Raw Data Ridge pada link di bawah ini :

LINK

Case Method : Ridge in R

Running Data Ridge pada link di bawah ini :

LINK

Case Method : Ridge in Python

Running Data Ridge pada link di bawah ini :

LINK

SEE YOU NEXT WEEK !

Ferdian Bangkit Wijaya, S.Stat., M.Si
NIP. 199005202024061001
ferdian.bangkit@untirta.ac.id