

• $\frac{\partial J_{OLS}(\beta)}{\partial \beta} = 2X^T X \beta$
Então:
 $\frac{\partial J_{OLS}(\beta)}{\partial \beta} = -2X^T y + 2X^T X \beta$
Igualando a zero para encontrar $\hat{\beta}_{OLS}$:
 $-2X^T y + 2X^T X \beta = 0$
 $2X^T X \beta = 2X^T y$
 $X^T X \beta = X^T y$
 $\beta_{OLS} = (X^T X)^{-1} X^T y$ (Se $X^T X$ for invertível)

2. Regularização L2 (Ridge Regression)

A Ridge Regression adiciona um termo de penalidade L2 na função de custo OLS. O termo de penalidade é $\lambda \sum_{j=1}^p \beta_j^2$, onde λ é o hiperparâmetro de regularização.

- **Função de Custo Ridge:**
 $J_{Ridge}(\beta) = \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^p \beta_j^2$

Em notação matricial, e para simplificar, vamos assumir que β inclui β_0 e que estamos regularizando todos os termos (incluindo β_0) com a matriz identidade I . Na prática, o β_0 geralmente não é regularizado, e a matriz I teria a primeira entrada $I_{0,0} = 0$.

$$J_{Ridge}(\beta) = (y - X\beta)^T (y - X\beta) + \lambda \beta^T \beta$$

Já sabemos que $(y - X\beta)^T (y - X\beta) = y^T y - 2y^T X\beta + \beta^T X^T X \beta$.

$$\beta^T \beta = \sum_{j=0}^p \beta_j^2$$

Então:

$$J_{Ridge}(\beta) = y^T y - 2y^T X\beta + \beta^T X^T X \beta + \lambda \beta^T \beta$$

Derivada da Função de Custo Ridge em relação a β :

Derivamos cada termo em relação a β . Já temos a derivada da parte OLS. Agora precisamos derivar o termo de penalidade.

$$\frac{\partial J_{Ridge}(\beta)}{\partial \beta} = \frac{\partial}{\partial \beta} (y^T y - 2y^T X\beta + \beta^T X^T X \beta + \lambda \beta^T \beta)$$

• A derivada da parte OLS é $-2X^T y + 2X^T X \beta$.

• Para o termo de penalidade: $\frac{\partial \lambda \beta^T \beta}{\partial \beta} = \lambda \frac{\partial \beta^T \beta}{\partial \beta} = \lambda (2\beta) = 2\lambda \beta$ (onde I é a matriz identidade).

Então:

$$\frac{\partial J_{Ridge}(\beta)}{\partial \beta} = -2X^T y + 2X^T X \beta + 2\lambda I \beta$$

Igualando a zero para encontrar $\hat{\beta}_{Ridge}$:

$$-2X^T y + 2X^T X \beta + 2\lambda I \beta = 0$$

Dividindo por 2:

$$-X^T y + X^T X \beta + \lambda I \beta = 0$$

Rearanjando os termos para isolar β :

$$X^T X \beta + \lambda I \beta = X^T y$$

Fatorando β :

$$(X^T X + \lambda I) \beta = X^T y$$

Finalmente, a solução para os coeficientes da Ridge Regression é:

$$\hat{\beta}_{Ridge} = (X^T X + \lambda I)^{-1} X^T y$$

Esta é a fórmula fundamental para calcular os coeficientes de uma regressão Ridge. A beleza dessa fórmula é que o termo λI adicionado a $X^T X$ torna a matriz invertível (não singular), mesmo em casos de multicolinearidade onde $X^T X$ sozinho seria singular.

User
da para vc traduzi com numeros nos cálculo

Model
Sim, claro! Vamos traduzir as funções de custo e suas derivadas para um exemplo numérico concreto.

Exemplo de Dados:

Vamos usar um conjunto de dados muito pequeno para tornar os cálculos manuais gerenciáveis.

- **Variável de destino (y):** [3, 5, 7]
- **Característica (x):** [1, 2, 3]

Assumimos um modelo linear simples com intercepto: $\hat{y} = \beta_0 + \beta_1 x$.

Matrizes:

Para o cálculo, precisamos da matriz X com uma coluna de 1's para o intercepto e o vetor y .

$$X = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}$$

$$y = \begin{bmatrix} 3 \\ 5 \\ 7 \end{bmatrix}$$

Vamos escolher um valor para λ . Seja $\lambda = 0.1$.

1. Regressão Linear por Mínimos Quadrados Ordinários (OLS) - Cálculos Numéricos

- **Objetivo:** Encontrar $\beta = [\beta_0, \beta_1]$ que minimize $J_{OLS}(\beta) = (y - X\beta)^T (y - X\beta)$.

Passo 1: Calcular $X^T X$

$$X^T = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix}$$

$$X^T X = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} (1 \cdot 1 + 1 \cdot 1 + 1 \cdot 1) & (1 \cdot 1 + 2 \cdot 1 + 3 \cdot 1) \\ (1 \cdot 1 + 2 \cdot 2 + 3 \cdot 3) & (1 \cdot 1 + 2 \cdot 3 + 3 \cdot 3) \end{bmatrix} = \begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix}$$

Passo 2: Calcular $X^T y$

$$X^T y = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 3 \\ 5 \\ 7 \end{bmatrix} = \begin{bmatrix} (1 \cdot 1 + 2 \cdot 1 + 3 \cdot 1) \\ (1 \cdot 2 + 2 \cdot 2 + 3 \cdot 3) \end{bmatrix} = \begin{bmatrix} 15 \\ 34 \end{bmatrix}$$

Passo 3: Calcular $(X^T X)^{-1}$

Para uma matriz $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, a inversa é $A^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$.

Determinante de $X^T X = (3 \cdot 14) - (6 \cdot 6) = 42 - 36 = 6$.

$$(X^T X)^{-1} = \frac{1}{6} \begin{bmatrix} 14 & -6 \\ -6 & 3 \end{bmatrix} = \begin{bmatrix} 14/6 & -6/6 \\ -6/6 & 3/6 \end{bmatrix} = \begin{bmatrix} 7/3 & -1 \\ -1 & 1/2 \end{bmatrix} \approx \begin{bmatrix} 2.333 & -0.0 \\ -1.0 & 0.5 \end{bmatrix}$$

Passo 4: Calcular $\hat{\beta}_{OLS} = (X^T X)^{-1} X^T y$

$$\hat{\beta}_{OLS} = \begin{bmatrix} 7/3 & -1 \\ -1 & 1/2 \end{bmatrix} \begin{bmatrix} 15 \\ 34 \end{bmatrix} = \begin{bmatrix} (7/3 \cdot 15 - 1 \cdot 34) \\ (-1 \cdot 15 + 1/2 \cdot 34) \end{bmatrix} = \begin{bmatrix} (35 - 34) \\ (-15 + 17) \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

Então, os coeficientes OLS são $\hat{\beta}_0 = 1$ e $\hat{\beta}_1 = 2$.

O modelo OLS é $\hat{y} = 1 + 2x$.

2. Regularização L2 (Ridge Regression) - Cálculos Numéricos

- **Objetivo:** Encontrar $\beta = [\beta_0, \beta_1]$ que minimize $J_{Ridge}(\beta) = (y - X\beta)^T (y - X\beta) + \lambda \beta^T \beta$.

Com $\lambda = 0.1$.

Fórmula: $\hat{\beta}_{Ridge} = (X^T X + \lambda I)^{-1} X^T y$

Passo 1: Calcular $X^T X$

Já calculamos: $X^T X = \begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix}$

Passo 2: Calcular λI

A matriz identidade I deve ter a mesma dimensão que $X^T X$, que é 2×2 .

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\lambda I = 0.1 \cdot \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix}$$

Passo 3: Calcular $(X^T X + \lambda I)$

$$(X^T X + \lambda I) = \begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix} + \begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix} = \begin{bmatrix} (3 + 0.1) & (6 + 0) \\ (6 + 0) & (14 + 0.1) \end{bmatrix} = \begin{bmatrix} 3.1 & 6 \\ 6 & 14.1 \end{bmatrix}$$

Passo 4: Calcular $(X^T X + \lambda I)^{-1}$

Determinante de $(X^T X + \lambda I) = (3.1 \cdot 14.1) - (6 \cdot 6) = 43.71 - 36 = 7.71$.

$$(X^T X + \lambda I)^{-1} = \frac{1}{7.71} \begin{bmatrix} 14.1 & -6 \\ -6 & 3 \end{bmatrix}$$

$$\approx \begin{bmatrix} 14.1/7.71 & -6/7.71 \\ -6/7.71 & 3/7.71 \end{bmatrix} \approx \begin{bmatrix} 1.8288 & -0.7782 \\ -0.7782 & 0.4021 \end{bmatrix}$$

Passo 5: Calcular $X^T y$

Já calculamos: $X^T y = \begin{bmatrix} 15 \\ 34 \end{bmatrix}$

Passo 6: Calcular $\hat{\beta}_{Ridge} = (X^T X + \lambda I)^{-1} X^T y$

$$\hat{\beta}_{Ridge} \approx \begin{bmatrix} 1.8288 & -0.7782 \\ -0.7782 & 0.4021 \end{bmatrix} \begin{bmatrix} 15 \\ 34 \end{bmatrix}$$

$$\hat{\beta}_{Ridge} \approx \begin{bmatrix} (1.8288 \cdot 15 - 0.7782 \cdot 34) \\ (1.8288 \cdot 34 - 0.7782 \cdot 15) \end{bmatrix}$$

$$\hat{\beta}_{Ridge} \approx \begin{bmatrix} (27.432 - 26.4588) \\ (-11.673 + 13.6714) \end{bmatrix}$$

$$\hat{\beta}_{Ridge} \approx \begin{bmatrix} 0.9732 \\ 1.0984 \end{bmatrix}$$

Então, os coeficientes Ridge são aproximadamente $\hat{\beta}_0 \approx 0.9732$ e $\hat{\beta}_1 \approx 1.0984$.

Comparação e Conclusão

- **Coeficientes OLS:** $\beta_0 = 1$, $\beta_1 = 2$
- **Coeficientes Ridge:** $\hat{\beta}_0 \approx 0.9732$, $\hat{\beta}_1 \approx 1.0984$

Neste exemplo específico, com dados perfeitos e sem multicolinearidade severa, a diferença entre OLS e Ridge é pequena. No entanto, você pode ver que os coeficientes Ridge foram ligeiramente "encolhidos" (tornaram-se um pouco menores em magnitude) em comparação com os OLS.

- β_0 (OLS: 1) para (Ridge: 0.9732) - Encolheu um pouco.
- β_1 (OLS: 2) para (Ridge: 1.0984) - Encolheu um pouco.

Se houvesse mais características ou se o valor de λ fosse maior, o efeito de encolhimento seria mais pronunciado. Em cenários de multicolinearidade ou overfitting, essa pequena diferença pode ser crucial para a estabilidade e capacidade de generalização do modelo.

Este exemplo numérico demonstra como a fórmula da Ridge Regression é aplicada usando operações matriciais, levando à obtenção dos coeficientes regularizados.

