# Regularizations

Ngoc Hoang Luong

University of Information Technology (UIT), VNU-HCM

April 13, 2023





# Chính quy hóa

Lương Ngọc Hoàng

Trường Đại học Công nghệ Thông tin (UIT), ĐHQG-HCM

Ngày 13 tháng 4 năm 2023

# Linear Regression

- In **linear regression**, the overall error function $E()$ is the mean squared error (MSE).





# Hồi quy tuyến tính

- Trong **hồi quy tuyến tính,** hàm lỗi tổng thể E() là lỗi bình phương trung bình (MSE).

# Linear Regression

- In **linear regression**, the overall error function $E()$ is the mean squared error (MSE).
- From the perspective of the parameters (i.e., the regression coefficients), we denote the error function as $E(\mathbf{b})$.

$$E(\mathbf{b}) = \frac{1}{n}(\mathbf{X}\mathbf{b} - \mathbf{y})^\top(\mathbf{X}\mathbf{b} - \mathbf{y})$$





# Hồi quy tuyến tính

- Trong hồi quy tuyến tính, hàm lỗi tổng thể E() là lỗi bình phương trung bình (MSE).
- Từ quan điểm của các tham số (nghĩa là các hệ số hồi quy), chúng ta ký hiệu hàm sai số là E(b).

$$E(b) = \frac{1}{N}(Xb \quad y) \quad (Xb \quad y)$$

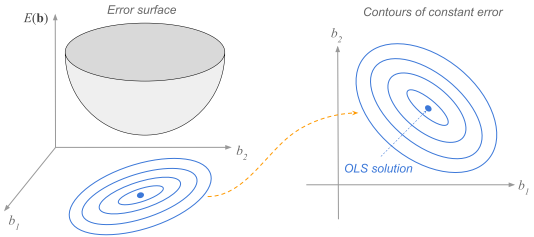
VIET NAM NATIONAL UNIVERSITY HCMC
UNIVERSITY OF INFORMATION TECHNOLOGY
VNUHCM - UIT
2 / 19

# Linear Regression

- In **linear regression**, the overall error function $E()$ is the mean squared error (MSE).
- From the perspective of the parameters (i.e., the regression coefficients), we denote the error function as $E(\mathbf{b})$.

$$E(\mathbf{b}) = \frac{1}{n}(\mathbf{X}\mathbf{b} - \mathbf{y})^\top(\mathbf{X}\mathbf{b} - \mathbf{y})$$

- Let's consider two inputs $X_1$ and $X_2$, and their corresponding parameters $b_1$ and $b_2$. The error function $E(\mathbf{b})$ generates a convex error surface with the shape of a bowl (or a paraboloid).







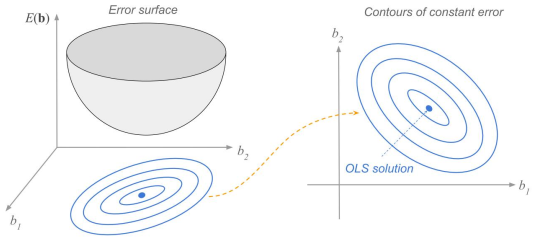

# Hồi quy tuyến tính •

Trong **hồi quy tuyến tính,** hàm lỗi tổng thể E() là lỗi bình phương trung bình (MSE).

- Từ quan điểm của các tham số (nghĩa là các hệ số hồi quy), chúng ta ký hiệu hàm sai số là E(b).

$$E(b) = \frac{1}{N}(Xb\quad y)\qquad(Xb\quad y)$$

- Hãy xem xét hai đầu vào X1 và X2 và tương ứng của chúng thông số b1 và b2. Hàm lỗi E(b) tạo ra một bề mặt lỗi lồi với hình dạng của một cái bát (hoặc một paraboloid).



# Linear Regression

$$E(\mathbf{b}) = \frac{1}{n}(\mathbf{X}\mathbf{b} - \mathbf{y})^{\top}(\mathbf{X}\mathbf{b} - \mathbf{y})$$

- In ordinary least squares (OLS), we minimize $E(\mathbf{b})$ **unconditionally**: without any restriction.





# hồi quy tuyến tính

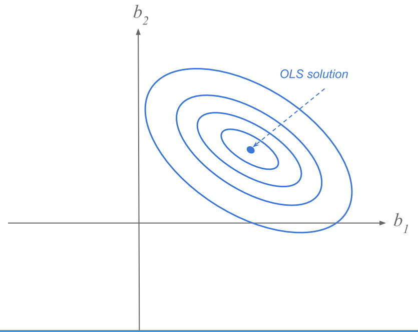
$$E(b) = \frac{1}{N}(Xb \quad y) \quad (Xb \quad y)$$

- Trong bình phương nhỏ nhất thông thường (OLS), chúng ta tối thiểu hóa E(b) một cách vô điều kiện: không có bất kỳ hạn chế nào.

# Linear Regression

$$E(\mathbf{b}) = \frac{1}{n}(\mathbf{X}\mathbf{b} - \mathbf{y})^\top(\mathbf{X}\mathbf{b} - \mathbf{y})$$

- In ordinary least squares (OLS), we minimize $E(\mathbf{b})$ **unconditionally**: without any restriction.
- The solution is indicated with a blue dot at the center of the elliptical contours of constant error.



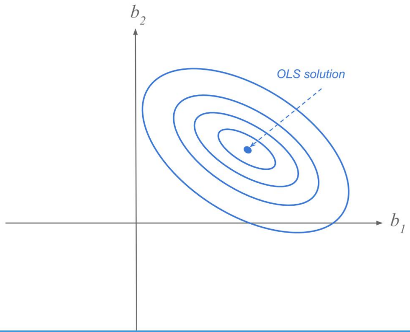





# hồi quy tuyến tính

$$E(b) = \frac{1}{N}(Xb \quad y) \qquad (Xb \quad y)$$

- Trong bình phương nhỏ nhất thông thường (OLS), chúng ta tối thiểu hóa E(b) một cách vô điều kiện: không có bất kỳ hạn chế nào.
- Dung dịch được biểu thị bằng một chấm màu xanh lam ở giữa đường viền elip của lỗi liên tục.



# Linear Regression - Constraining Regression Coefficients

$$E(\mathbf{b}) = \frac{1}{n}(\mathbf{X}\mathbf{b} - \mathbf{y})^\top(\mathbf{X}\mathbf{b} - \mathbf{y})$$

- We would like now to impose a restriction on the squared magnitude of the regression coefficients.





# Hồi quy tuyến tính - Hệ số hồi quy ràng buộc

$$E(b) = \frac{1}{N}(Xb \quad y) \quad (Xb \quad y)$$

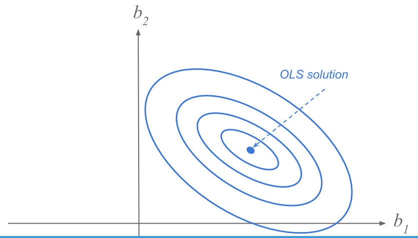
- Bây giờ chúng tôi muốn áp đặt một hạn chế đối với bình phương độ lớn của các hệ số hồi quy.

# Linear Regression - Constraining Regression Coefficients

$$E(\mathbf{b}) = \frac{1}{n}(\mathbf{X}\mathbf{b} - \mathbf{y})^\top(\mathbf{X}\mathbf{b} - \mathbf{y})$$

- We would like now to impose a restriction on the squared magnitude of the regression coefficients.
- We still minimize $E(\mathbf{b})$, but now we require the following condition on $b_1, b_2, \ldots, b_p$:

$$\sum_{j=1}^{p} b_j^2 \leq c \qquad (1)$$





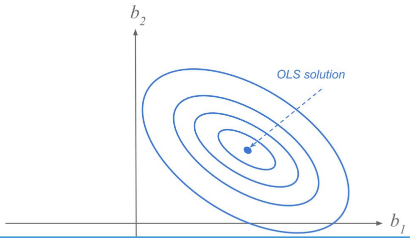

Machine Translated by Google

# Hồi quy tuyến tính - Hệ số hồi quy ràng buộc

E(b) = 1/N (Xb y) (Xb y)

- Bây giờ chúng tôi muốn áp đặt một hạn chế đối với bình phương độ lớn của các hệ số hồi quy.
- Ta vẫn giảm thiểu E(b), nhưng bây giờ ta cần điều kiện sau trên b1, b2, . . . , bp:

P
2 b ≤ cj
j=1 (1)



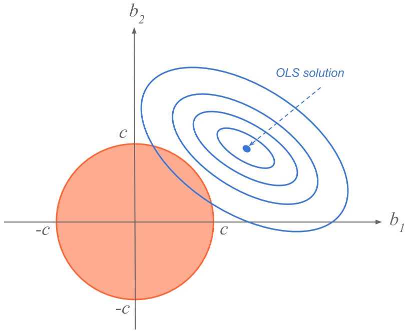

VIET NAM NATIONAL UNIVERSITY HCMC
UNIVERSITY OF INFORMATION TECHNOLOGY
VNUHCM - UIT
4 / 19

# Linear Regression - Constraining Regression Coefficients

$$E(\mathbf{b}) = \frac{1}{n}(\mathbf{Xb}-\mathbf{y})^\top(\mathbf{Xb}-\mathbf{y})$$

- We have a constrained minimization of $E(\mathbf{b})$ for some "budget" $c$:

$$\min_{\mathbf{b}} \left\{ \frac{1}{n}(\mathbf{Xb}-\mathbf{y})^\top(\mathbf{Xb}-\mathbf{y}) \right\} \text{ st } \|\mathbf{b}\|_2^2 = \mathbf{b}^\top\mathbf{b} \leq c$$





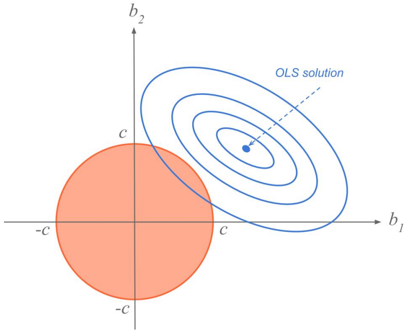



# Hồi quy tuyến tính - Hệ số hồi quy ràng buộc

$$E(b) = \frac{1}{N}(Xb \quad y) \quad (Xb \quad y)$$

- Chúng tôi có mức tối thiểu hóa hạn chế của E(b) đối với một số "ngân sách" c:

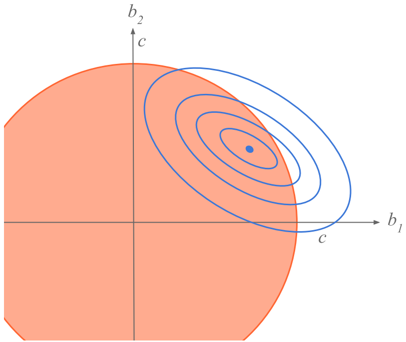
tối thiểu b { $\frac{1}{N}$ (Xb y) (Xb y)} st b $^2_2$ = b b ≤ c



# Linear Regression - Constraining Regression Coefficients

$$\min_{\mathbf{b}} \left\{ \frac{1}{n} (\mathbf{X}\mathbf{b} - \mathbf{y})^\top (\mathbf{X}\mathbf{b} - \mathbf{y}) \right\} \text{ st } \|\mathbf{b}\|_2^2 = \mathbf{b}^\top \mathbf{b} \leq c$$

- If we choose too big values of $c$, we could have a big enough constraint that includes the OLS solution.



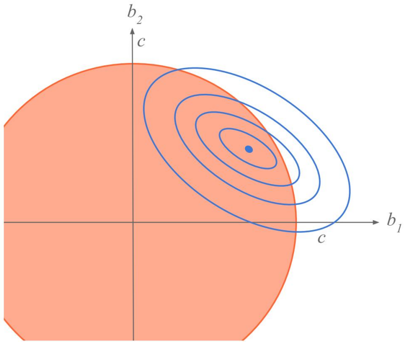





# Hồi quy tuyến tính - Hệ số hồi quy ràng buộc

tối thiểu b { $\frac{1}{N}$ (Xb y) (Xb y)} st b $^2_2$ = b b ≤ c

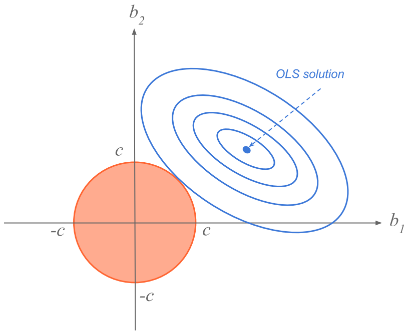
- Nếu chúng ta chọn giá trị c quá lớn, chúng ta có thể có một ràng buộc bao gồm giải pháp OLS.



# Linear Regression - Constraining Regression Coefficients

$$\min_{\mathbf{b}} \left\{ \frac{1}{n} (\mathbf{X}\mathbf{b} - \mathbf{y})^{\mathsf{T}} (\mathbf{X}\mathbf{b} - \mathbf{y}) \right\} \text{ st } \|\mathbf{b}\|_2^2 = \mathbf{b}^{\mathsf{T}}\mathbf{b} \leq c$$

- We could make the budget stricter by reducing the value of $c$



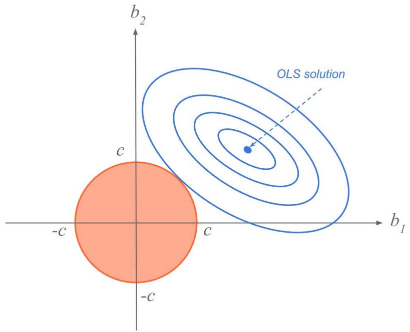





# Hồi quy tuyến tính - Hệ số hồi quy ràng buộc

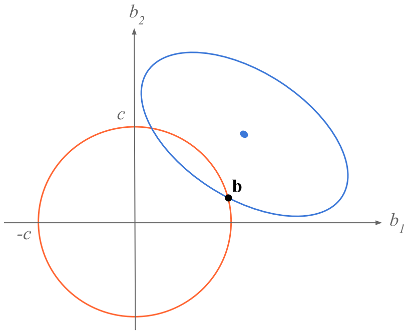
tối thiểu b { 1/N (Xb y) (Xb y)} st b 2/2 = b b ≤ c

- Chúng ta có thể thắt chặt ngân sách hơn bằng cách giảm giá trị của c



# Linear Regression - A New Minimization Solution

- Let's consider one elliptical contour of constant error, a given budget $c$, and a point $\mathbf{b}$ satisfying the budget constraint

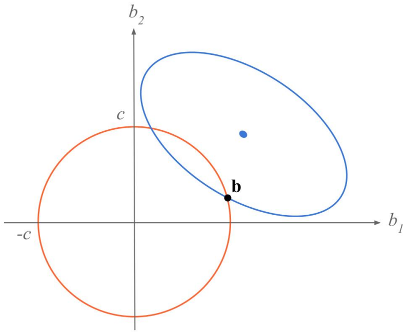






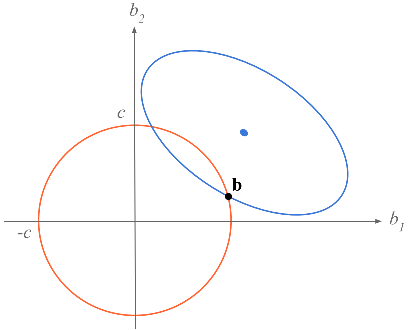

# Hồi quy tuyến tính - Một giải pháp tối thiểu hóa mới

- Hãy xem xét một đường bao hình elip có sai số không đổi, ngân sách c cho trước và điểm b thỏa mãn giới hạn ngân sách



# Linear Regression - A New Minimization Solution

- Let's consider one elliptical contour of constant error, a given budget $c$, and a point $\mathbf{b}$ satisfying the budget constraint



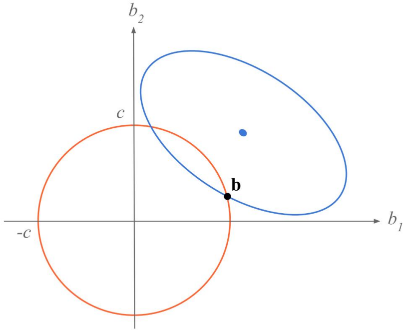

- $\mathbf{b}^\top\mathbf{b} = c$. However, this point does not fully minimize $E(\mathbf{b})$.





# Hồi quy tuyến tính - Một giải pháp tối thiểu hóa mới

- Hãy xem xét một đường bao hình elip có sai số không đổi, ngân sách c cho trước và điểm b thỏa mãn giới hạn ngân sách

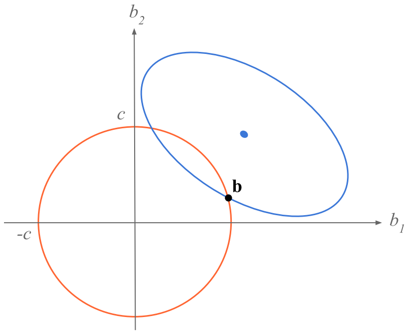



- b b = c. Tuy nhiên, điểm này không giảm thiểu hoàn toàn E(b).

VIET NAM NATIONAL UNIVERSITY HCMC
UNIVERSITY OF INFORMATION TECHNOLOGY
VNUHCM - UIT
8 / 19

# Linear Regression - A New Minimization Solution

- Let's consider one elliptical contour of constant error, a given budget $c$, and a point $\mathbf{b}$ satisfying the budget constraint



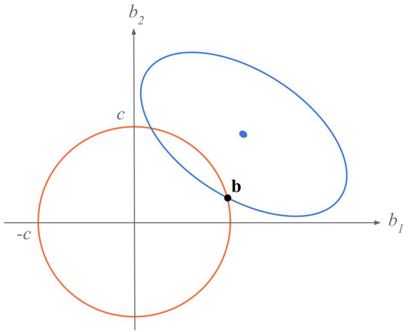

- $\mathbf{b}^\top\mathbf{b} = c$. However, this point does not fully minimize $E(\mathbf{b})$.
- We could still find other $\mathbf{b}$ along the circle that would give us smaller $E(\mathbf{b})$.





# Hồi quy tuyến tính - Một giải pháp tối thiểu hóa mới

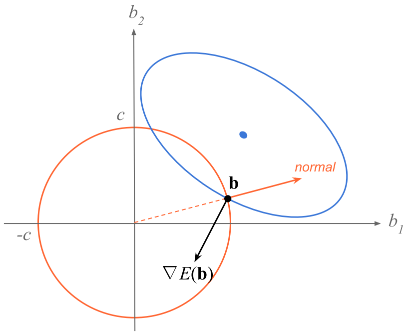
- Hãy xem xét một đường bao hình elip có sai số không đổi, ngân sách c cho trước và điểm b thỏa mãn giới hạn ngân sách




- b b = c. Tuy nhiên, điểm này không giảm thiểu hoàn toàn E(b). •

Chúng tôi vẫn có thể tìm thấy các b khác dọc theo vòng tròn sẽ cho chúng tôi E(b) nhỏ hơn.

# Linear Regression - A New Minimization Solution

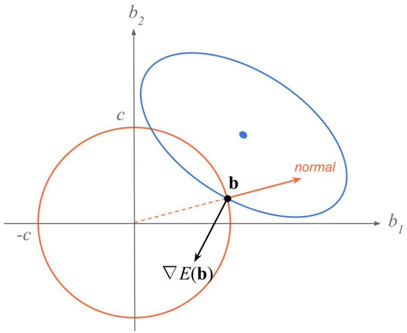



- The gradient $\nabla E(\mathbf{b})$ points in the direction orthogonal to the contour ellipse, i.e., the direction of largest change of $E(\mathbf{b})$.



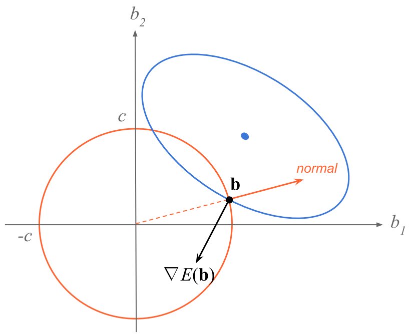


# Hồi quy tuyến tính - Một giải pháp tối thiểu hóa mới




- Độ dốc E(b) chỉ theo hướng trực giao với đường bao elip, tức là hướng thay đổi lớn nhất của E(b).

# Linear Regression - A New Minimization Solution

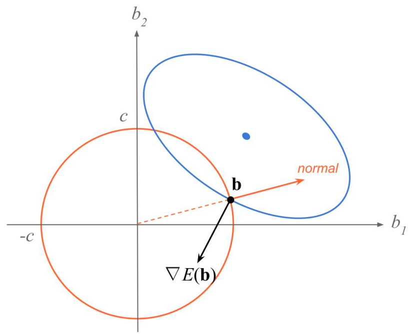



- The gradient $\nabla E(\mathbf{b})$ points in the direction orthogonal to the contour ellipse, i.e., the direction of largest change of $E(\mathbf{b})$.
- The direction of $\mathbf{b}$ is orthogonal to the circumference of the constraint (normal vector). The angle between the gradient and the normal vector is less than 180 degrees. We can find better $\mathbf{b}$ points that make the error smaller. Where is that optimal $\mathbf{b}^*$?





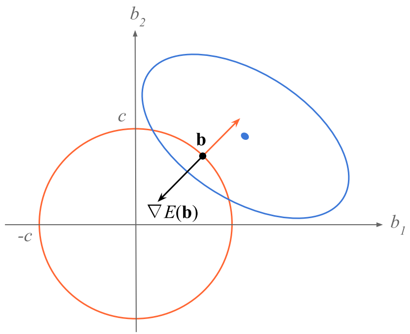
# Hồi quy tuyến tính - Một giải pháp tối thiểu hóa mới




• Độ dốc E(b) chỉ theo hướng trực giao với đường bao elip, tức là hướng thay đổi lớn nhất của E(b). • Hướng của b là trực giao với chu vi của ràng buộc (vectơ pháp tuyến). Góc giữa gradient và vector pháp tuyến nhỏ hơn 180 độ. Chúng ta có thể tìm thấy điểm b tốt hơn làm cho sai số nhỏ hơn. Đó là tối ưu ở đâu b ?


VIET NAM NATIONAL UNIVERSITY HCMC
UNIVERSITY OF INFORMATION TECHNOLOGY
VNUHCM - UIT
9 / 19

# Linear Regression - A New Minimization Solution

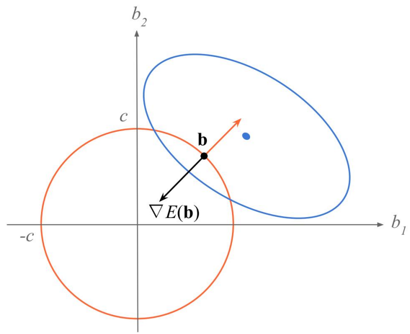



- The optimal vector $\mathbf{b}^*$ corresponds to the one that is exactly the opposite of $\nabla E(\mathbf{b})$. The gradient and the normal vectors are anti-parallel: $\nabla E(\mathbf{b}^*) \propto -\mathbf{b}^*$.



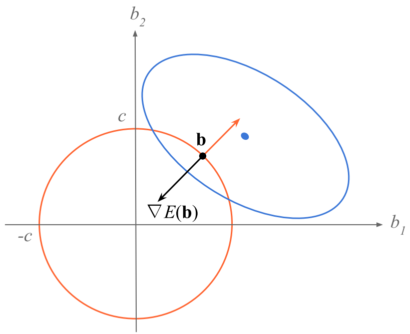


# Hồi quy tuyến tính - Một giải pháp tối thiểu hóa mới




- Vectơ b tối ưu đối của tương ứng với một trong đó là chính xác E(b). Độ dốc và các vectơ pháp tuyến là phản song song: E(b ) b .

# Linear Regression - A New Minimization Solution



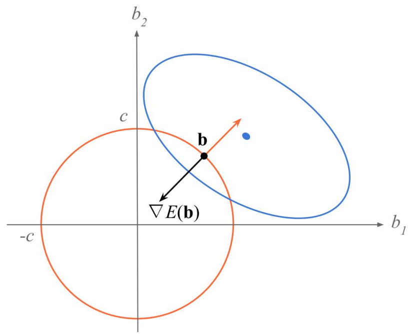

- The optimal vector $\mathbf{b}^*$ corresponds to the one that is exactly the opposite of $\nabla E(\mathbf{b})$. The gradient and the normal vectors are anti-parallel: $\nabla E(\mathbf{b}^*) \propto -\mathbf{b}^*$.
- We choose a proportionality constant of $-2(\lambda/n)$

$$\nabla E(\mathbf{b}^*) = -2\frac{\lambda}{n}\mathbf{b}^*$$





# Hồi quy tuyến tính - Một giải pháp tối thiểu hóa mới




- Vectơ b tối ưu đối tương ứng với một trong đó là chính xác của E(b). Độ dốc và các vectơ pháp tuyến là phản song song: E(b ) b .
- Chúng tôi chọn hằng số tỷ lệ là 2(λ/n)

E(b ) = 2 λ/N b

Eb λ + 2 = 0

# Linear Regression - A New Minimization Solution

$$\nabla E(\mathbf{b}^*) = -2\frac{\lambda}{n}\mathbf{b}^*$$

$$\nabla E(\mathbf{b}^*) + 2\frac{\lambda}{n}\mathbf{b}^* = \mathbf{0}$$

- The above expression is the gradient of the following function:

$$f(\mathbf{b}) = E(\mathbf{b}) + \frac{\lambda}{n}\mathbf{b}^\top\mathbf{b}$$

$$\nabla f(\mathbf{b}^*) = \nabla E(\mathbf{b}^*) + 2\frac{\lambda}{n}\mathbf{b}^*$$





# Hồi quy tuyến tính - Một giải pháp tối thiểu hóa mới

E(b λ ) = 2 b
N

E(b λ ) + 2 b = 0
N

- Biểu thức trên là gradient của hàm sau:

f(b) = E(b) + λ b b
N

f(b ) = E(b λ ) + 2 b
N

# Linear Regression - A New Minimization Solution

$$\nabla E(\mathbf{b}^*) = -2\frac{\lambda}{n}\mathbf{b}^*$$
$$\nabla E(\mathbf{b}^*) + 2\frac{\lambda}{n}\mathbf{b}^* = \mathbf{0}$$

- The above expression is the gradient of the following function:

$$f(\mathbf{b}) = E(\mathbf{b}) + \frac{\lambda}{n}\mathbf{b}^\top\mathbf{b}$$
$$\nabla f(\mathbf{b}^*) = \nabla E(\mathbf{b}^*) + 2\frac{\lambda}{n}\mathbf{b}^*$$

- Now, our minimization problem becomes:

$$\min_{\mathbf{b}}\left\{E(\mathbf{b}) + \frac{\lambda}{n}\mathbf{b}^\top\mathbf{b}\right\}$$





# Hồi quy tuyến tính - Một giải pháp tối thiểu hóa mới

E(b λ ) = 2 b / N

E(b λ ) + 2 b / N = 0

- Biểu thức trên là gradient của hàm sau:

f(b) = E(b) + λ b b / N

f(b ) = E(b λ ) + 2 b / N

- Bây giờ, bài toán tối thiểu hóa của chúng ta trở thành:

tối thiểu b {E(b) + λ / N b b}

# Linear Regression - A New Minimization Solution

$$
\begin{aligned}
f(\mathbf{b}) &= E(\mathbf{b}) + \frac{\lambda}{n}\mathbf{b}^\top\mathbf{b} = \frac{1}{n}(\mathbf{X}\mathbf{b}-\mathbf{y})^\top(\mathbf{X}\mathbf{b}-\mathbf{y}) + \frac{\lambda}{n}\mathbf{b}^\top\mathbf{b} \\
&= \frac{1}{n}\mathbf{b}^\top\mathbf{X}^\top\mathbf{X}\mathbf{b} - \frac{2}{n}\mathbf{b}^\top\mathbf{X}^\top\mathbf{y} + \frac{1}{n}\mathbf{y}^\top\mathbf{y} + \frac{\lambda}{n}\mathbf{b}^\top\mathbf{b}
\end{aligned}
$$

- Compute the gradient:

$$
\nabla f(\mathbf{b}) = \frac{2}{n}\mathbf{X}^\top\mathbf{X}\mathbf{b} - \frac{2}{n}\mathbf{X}^\top\mathbf{y} + \frac{2\lambda}{n}\mathbf{b}
$$





# Hồi quy tuyến tính - Một giải pháp tối thiểu hóa mới

$$
\begin{aligned}
f(b) &= E(b) + \frac{\lambda}{N} b \quad b = \frac{1}{N}(Xb \quad y) \quad (Xb \quad y) + \frac{\lambda}{N} b \quad b \\
&= \frac{1}{N}\ 2\ b \quad X \quad Xb \quad \frac{}{N} b \quad X \quad y + \frac{1}{N} \lambda\ y \quad y + \frac{b}{N} \quad b
\end{aligned}
$$

- Tính độ dốc:

$$
f(b) = \frac{2\ 2}{N} X \quad Xb \quad \frac{}{N} X \quad y + \frac{2\lambda}{N} b
$$

# Linear Regression - A New Minimization Solution

$$f(\mathbf{b}) = E(\mathbf{b}) + \frac{\lambda}{n}\mathbf{b}^\top\mathbf{b} = \frac{1}{n}(\mathbf{X}\mathbf{b}-\mathbf{y})^\top(\mathbf{X}\mathbf{b}-\mathbf{y}) + \frac{\lambda}{n}\mathbf{b}^\top\mathbf{b}$$

$$= \frac{1}{n}\mathbf{b}^\top\mathbf{X}^\top\mathbf{X}\mathbf{b} - \frac{2}{n}\mathbf{b}^\top\mathbf{X}^\top\mathbf{y} + \frac{1}{n}\mathbf{y}^\top\mathbf{y} + \frac{\lambda}{n}\mathbf{b}^\top\mathbf{b}$$

- Compute the gradient:

$$\nabla f(\mathbf{b}) = \frac{2}{n}\mathbf{X}^\top\mathbf{X}\mathbf{b} - \frac{2}{n}\mathbf{X}^\top\mathbf{y} + \frac{2\lambda}{n}\mathbf{b}$$

- Setting $\nabla f(\mathbf{b})$ to $\mathbf{0}$, and solve for the **ridge regression** of $\mathbf{b}$:

$$\mathbf{X}^\top\mathbf{X}\mathbf{b} - \mathbf{X}^\top\mathbf{y} + \lambda\mathbf{b} = \mathbf{0}$$

$$\mathbf{b}(\mathbf{X}^\top\mathbf{X} + \lambda\mathbf{I}) = \mathbf{X}^\top\mathbf{y}$$

$$\mathbf{b}_{RR} = (\mathbf{X}^\top\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}^\top\mathbf{y}$$





# Hồi quy tuyến tính - Một giải pháp tối thiểu hóa mới

$$f(b) = E(b) + \frac{\lambda}{N} b \quad b = \frac{1}{N}(Xb \quad y) \quad (Xb \quad y) + \frac{\lambda}{N} b \quad b$$

$$= \frac{1}{N} \; 2 \; b \quad X \quad Xb \quad \frac{}{N} b \quad X \quad y + \frac{1}{N} \lambda \; y \quad y + \frac{}{N} b \quad b$$

- Tính độ dốc:

$$f(b) = \frac{2\;2}{N} X \quad Xb \quad \frac{}{N} X \quad y + \frac{2\lambda}{N} b$$

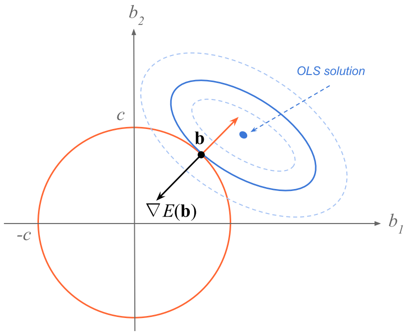
- Đặt  f(b) thành 0 và giải hồi quy sườn của b:

$$X \quad Xb \quad X \quad y + \lambda b =$$

$$0 \; b(X \quad X + \lambda I) =$$

$$X \quad y \; bRR = (X \quad X + \lambda I) \quad 1X \quad y$$

# Ridge Regression



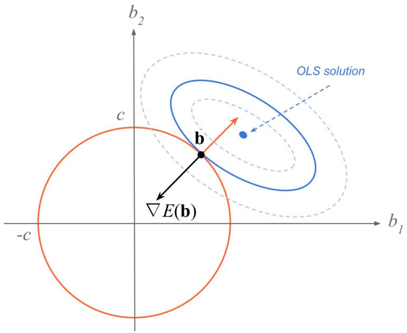

ridge coefficients $\mathbf{b}_{RR} = (\mathbf{X}^{\top}\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}^{\top}\mathbf{y}$

- The optimal vector $\mathbf{b}$ that minimizes $E(\mathbf{b})$ and satisfies the constraint will be on a contour of constant error.





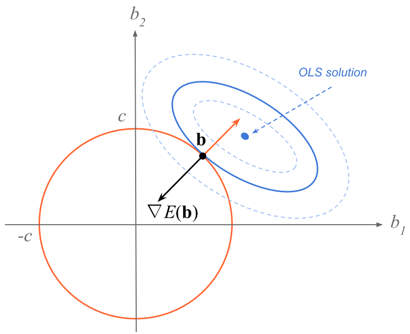
# Hồi quy sườn




hệ số sườn bRR = (X X + λI) 1X y

- Vectơ b tối ưu làm cực tiểu E(b) và thỏa mãn ràng buộc sẽ nằm trên một đường bao có sai số không đổi.

# Ridge Regression



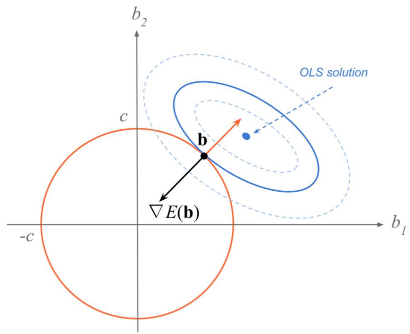

ridge coefficients $\mathbf{b}_{RR} = (\mathbf{X}^\top\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}^\top\mathbf{y}$

- The optimal vector $\mathbf{b}$ that minimizes $E(\mathbf{b})$ and satisfies the constraint will be on a contour of constant error.
- The above illustration shows that the direction of the optimal vector $\mathbf{b}$ does not point in the direction of the OLS solution.





# Hồi quy sườn




hệ số sườn bRR = (X X + λI) 1X y

- Vectơ b tối ưu làm cực tiểu E(b) và thỏa mãn ràng buộc sẽ nằm trên một đường bao có sai số không đổi.
- Hình minh họa trên cho thấy hướng của vectơ b không chỉ theo hướng của giải pháp OLS.

# Ridge Regression

- If we set the hyperparameter $\lambda = 0$, the ridge solution is the same as the OLS solution.

$$\mathbf{b}_{RR} = (\mathbf{X}^\top\mathbf{X} + 0\mathbf{I})^{-1}\mathbf{X}^\top\mathbf{y} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y} = \mathbf{b}_{OLS}$$





# Hồi quy sườn

- Nếu chúng ta đặt siêu tham số λ = 0, giải pháp sườn giống như giải pháp OLS.

bRR = (X X + 0I) 1X y = (X X) 1X y = bOLS

# Ridge Regression

- If we set the hyperparameter $\lambda = 0$, the ridge solution is the same as the OLS solution.

$$\mathbf{b}_{RR} = (\mathbf{X}^\top\mathbf{X} + 0\mathbf{I})^{-1}\mathbf{X}^\top\mathbf{y} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y} = \mathbf{b}_{OLS}$$

- If $\lambda$ becomes larger and larger, the terms on the diagonal of $(\mathbf{X}^\top\mathbf{X} + \lambda\mathbf{I})$ will be dominated by such $\lambda$ values.





# Hồi quy sườn

- Nếu chúng ta đặt siêu tham số λ = 0, giải pháp sườn giống như giải pháp OLS.

bRR = (X X + 0I) 1X y = (X X) 1X y = bOLS

- Nếu λ ngày càng lớn thì các số hạng trên đường chéo của (X X + λI) sẽ bị chi phối bởi các giá trị λ đó.

# Ridge Regression

- If we set the hyperparameter $\lambda = 0$, the ridge solution is the same as the OLS solution.

$$\mathbf{b}_{RR} = (\mathbf{X}^\top\mathbf{X} + 0\mathbf{I})^{-1}\mathbf{X}^\top\mathbf{y} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y} = \mathbf{b}_{OLS}$$

- If $\lambda$ becomes larger and larger, the terms on the diagonal of $(\mathbf{X}^\top\mathbf{X} + \lambda\mathbf{I})$ will be dominated by such $\lambda$ values.
- The matrix inverse, $(\mathbf{X}^\top\mathbf{X} + \lambda\mathbf{I})^{-1}$, will be dominated by the inverse of the terms on its diagonal:

$$\lambda \gg 0 \Rightarrow (\mathbf{X}^\top\mathbf{X} + \lambda\mathbf{I})^{-1} \to \frac{1}{\lambda}\mathbf{I}$$
$$\lambda \to \infty \Rightarrow (\mathbf{X}^\top\mathbf{X} + \lambda\mathbf{I})^{-1} \to \mathbf{0}_{(p,p)}$$
$$\lambda \to \infty \Rightarrow \mathbf{b}_{RR} = \mathbf{0}$$





# Hồi quy sườn

- Nếu chúng ta đặt siêu tham số λ = 0, giải pháp sườn giống như giải pháp OLS.

bRR = (X X + 0I) 1X y = (X X) 1X y = bOLS

- Nếu λ ngày càng lớn thì các số hạng trên đường chéo của (X X + λI) sẽ bị chi phối bởi các giá trị λ đó. sẽ bị chi
- Ma trận nghịch đảo, (X X + λI) 1 , phối bởi nghịch đảo của các điều khoản trên đường chéo của nó:

λ 0 (X X + λI) 1 1 / tôi

λ ∞ (X X + λI) λ 1 0(p,p)

∞ bRR = 0

# Ridge Regression - Relation between $\lambda$ and $c$

- Our two minimization problems:

$$\min_{\mathbf{b}} \left\{ \frac{1}{n} (\mathbf{X}\mathbf{b} - \mathbf{y})^{\top} (\mathbf{X}\mathbf{b} - \mathbf{y}) \right\} \text{ st } \|\mathbf{b}\|_2^2 = \mathbf{b}^{\top}\mathbf{b} \leq c$$

$$\min_{\mathbf{b}} \left\{ E(\mathbf{b}) + \frac{\lambda}{n} \mathbf{b}^{\top}\mathbf{b} \right\}$$





# Hồi quy sườn - Mối quan hệ giữa λ và c

- Hai vấn đề giảm thiểu của chúng tôi:

tối thiểu
b { 1/N (Xb y) (Xb y)} st b 2 2 = b b ≤ c

tối thiểu
b {E(b) + λ/N b b}

# Ridge Regression - Relation between $\lambda$ and $c$

- Our two minimization problems:

$$\min_{\mathbf{b}} \left\{ \frac{1}{n} (\mathbf{X}\mathbf{b} - \mathbf{y})^\top (\mathbf{X}\mathbf{b} - \mathbf{y}) \right\} \text{ st } \|\mathbf{b}\|_2^2 = \mathbf{b}^\top \mathbf{b} \le c$$

$$\min_{\mathbf{b}} \left\{ E(\mathbf{b}) + \frac{\lambda}{n} \mathbf{b}^\top \mathbf{b} \right\}$$

- Imposing a large budget constraint $c$ causes $\lambda$ to become smaller. The smaller the $\lambda$, the closer $\mathbf{b}_{RR}$ to the OLS solution.





# Hồi quy sườn - Mối quan hệ giữa λ và c

- Hai vấn đề giảm thiểu của chúng tôi:

tối thiểu b { 1 / N (Xb y) (Xb y)} st b 2 2 = b b ≤ c

tối thiểu b {E(b) + λ / N b b}

- Áp đặt một ràng buộc ngân sách lớn c làm cho λ trở nên nhỏ hơn. λ càng nhỏ, bRR càng gần nghiệm OLS.

# Ridge Regression - Relation between $\lambda$ and $c$

- Our two minimization problems:

$$\min_{\mathbf{b}} \left\{ \frac{1}{n} (\mathbf{Xb} - \mathbf{y})^{\top} (\mathbf{Xb} - \mathbf{y}) \right\} \text{ st } \|\mathbf{b}\|_2^2 = \mathbf{b}^{\top}\mathbf{b} \leq c$$

$$\min_{\mathbf{b}} \left\{ E(\mathbf{b}) + \frac{\lambda}{n} \mathbf{b}^{\top}\mathbf{b} \right\}$$

- Imposing a large budget constraint $c$ causes $\lambda$ to become smaller. The smaller the $\lambda$, the closer $\mathbf{b}_{RR}$ to the OLS solution.
- Imposing a small budget constraint $c$ causes $\lambda$ to become larger. The larger the $\lambda$, the smaller the ridge coefficients.

$$\uparrow c \Rightarrow \downarrow \lambda$$

$$\downarrow c \Rightarrow \uparrow \lambda$$





# Hồi quy sườn - Mối quan hệ giữa λ và c

- Hai vấn đề giảm thiểu của chúng tôi:

tối thiểu b { 1/N (Xb y) (Xb y)} st b 2 2 = b b ≤ c

tối thiểu b {E(b) + λ/N b b}

- Áp đặt một ràng buộc ngân sách lớn c làm cho λ trở nên nhỏ hơn. λ càng nhỏ, bRR càng gần nghiệm OLS.
- Áp đặt một ràng buộc ngân sách nhỏ c làm cho λ trở nên lớn hơn. λ càng lớn thì hệ số sườn núi càng nhỏ.

c

λ c λ

# Ridge Regression - How to find $\lambda$?

- In ridge regression, $\lambda$ is a hyperparameter that we need to tune.





# Hồi quy sườn - Cách tìm λ?

- Trong hồi quy sườn núi, λ là một siêu tham số mà chúng ta cần điều chỉnh.

# Ridge Regression - How to find $\lambda$?

- In ridge regression, $\lambda$ is a hyperparameter that we need to tune.
- We randomly split the training data:

$$\mathcal{D}_{train} = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)\}$$

into $K$ folds:

$$\mathcal{D}_{train} = \mathcal{D}_{fold-1} \cup \mathcal{D}_{fold-2} \cup \ldots \cup \mathcal{D}_{fold-K}$$





# Hồi quy sườn - Cách tìm λ?

- Trong hồi quy sườn núi, λ là một siêu tham số mà chúng ta cần điều chỉnh.
- Chúng tôi chia ngẫu nhiên dữ liệu huấn luyện:

Dtrain = {(x1, y1), . . . , (xn, yn)}

thành K nếp gấp:

Dtrain = Df old 1 Df old 2 . . . Df già K

# Ridge Regression - How to find $\lambda$?

- In ridge regression, $\lambda$ is a hyperparameter that we need to tune.
- We randomly split the training data:

$$\mathcal{D}_{train} = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)\}$$

into $K$ folds:

$$\mathcal{D}_{train} = \mathcal{D}_{fold-1} \cup \mathcal{D}_{fold-2} \cup \ldots \cup \mathcal{D}_{fold-K}$$

- Each fold set $\mathcal{D}_{fold-k}$ plays the role of an evaluation $\mathcal{D}_{eval-k}$. The corresponding $K$ training sets:





# Hồi quy sườn - Cách tìm λ?

- Trong hồi quy sườn núi, λ là một siêu tham số mà chúng ta cần điều chỉnh. •

Chúng tôi chia ngẫu nhiên dữ liệu huấn luyện:

Dtrain = {(x1, y1), . . . , (xn, yn)}

thành K nếp gấp:

Dtrain = Df old 1 Df old 2 . . . Df già K

- Mỗi tập hợp nếp gấp Df old k đóng vai trò của một Deval k đánh giá. Các tập huấn luyện K tương ứng:

VIET NAM NATIONAL UNIVERSITY HCMC
UNIVERSITY OF INFORMATION TECHNOLOGY
VNUHCM - UIT
16 / 19

# Ridge Regression - How to find $\lambda$?

- In ridge regression, $\lambda$ is a hyperparameter that we need to tune.
- We randomly split the training data:

$$\mathcal{D}_{train} = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)\}$$

into $K$ folds:

$$\mathcal{D}_{train} = \mathcal{D}_{fold-1} \cup \mathcal{D}_{fold-2} \cup \ldots \cup \mathcal{D}_{fold-K}$$

- Each fold set $\mathcal{D}_{fold-k}$ plays the role of an evaluation $\mathcal{D}_{eval-k}$. The corresponding $K$ training sets:
  - $\mathcal{D}_{train-1} = \mathcal{D}_{train} \setminus \mathcal{D}_{fold-1}$





# Hồi quy sườn - Cách tìm λ?

- Trong hồi quy sườn núi, λ là một siêu tham số mà chúng ta cần điều chỉnh. •

Chúng tôi chia ngẫu nhiên dữ liệu huấn luyện:

Dtrain = {(x1, y1), . . . , (xn, yn)}

thành K nếp gấp:

Dtrain = Df old 1 Df old 2 · · · Df già K

- Mỗi tập hợp nếp gấp Df old k đóng vai trò của một Deval k đánh giá. K tập huấn luyện tương ứng: • Dtrain 1 =

Dtrain / Dfold 1

VIET NAM NATIONAL UNIVERSITY HCMC
UNIVERSITY OF INFORMATION TECHNOLOGY
VNUHCM - UIT
16 / 19

# Ridge Regression - How to find $\lambda$?

- In ridge regression, $\lambda$ is a hyperparameter that we need to tune.
- We randomly split the training data:

$$\mathcal{D}_{train} = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)\}$$

into $K$ folds:

$$\mathcal{D}_{train} = \mathcal{D}_{fold-1} \cup \mathcal{D}_{fold-2} \cup \ldots \cup \mathcal{D}_{fold-K}$$

- Each fold set $\mathcal{D}_{fold-k}$ plays the role of an evaluation $\mathcal{D}_{eval-k}$. The corresponding $K$ training sets:
  - $\mathcal{D}_{train-1} = \mathcal{D}_{train} \setminus \mathcal{D}_{fold-1}$
  - $\mathcal{D}_{train-2} = \mathcal{D}_{train} \setminus \mathcal{D}_{fold-2}$





# Hồi quy sườn - Cách tìm λ?

- Trong hồi quy sườn núi, λ là một siêu tham số mà chúng ta cần điều chỉnh. •

Chúng tôi chia ngẫu nhiên dữ liệu huấn luyện:

Dtrain = {(x1, y1), . . . , (xn, yn)}

thành K nếp gấp:

Dtrain = Df old 1 Df old 2 . . . Df già K

- Mỗi tập hợp nếp gấp Df old k đóng vai trò của một Deval k đánh giá. K tập huấn luyện tương ứng: •

Dtrain 1 = Dtrain / Dfold 1

- Dtrain 2 = Dtrain / Dfold 2

VIET NAM NATIONAL UNIVERSITY HCMC
UNIVERSITY OF INFORMATION TECHNOLOGY
VNUHCM - UIT
16 / 19

# Ridge Regression - How to find $\lambda$?

- In ridge regression, $\lambda$ is a hyperparameter that we need to tune.
- We randomly split the training data:

$$\mathcal{D}_{train} = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)\}$$

into $K$ folds:

$$\mathcal{D}_{train} = \mathcal{D}_{fold-1} \cup \mathcal{D}_{fold-2} \cup \ldots \cup \mathcal{D}_{fold-K}$$

- Each fold set $\mathcal{D}_{fold-k}$ plays the role of an evaluation $\mathcal{D}_{eval-k}$. The corresponding $K$ training sets:
  - $\mathcal{D}_{train-1} = \mathcal{D}_{train} \setminus \mathcal{D}_{fold-1}$
  - $\mathcal{D}_{train-2} = \mathcal{D}_{train} \setminus \mathcal{D}_{fold-2}$
  - ...





# Hồi quy sườn - Cách tìm λ?

• Trong hồi quy sườn núi, λ là một siêu tham số mà chúng ta cần điều chỉnh. •

Chúng tôi chia ngẫu nhiên dữ liệu huấn luyện:

Dtrain = {(x1, y1), . . . , (xn, yn)}

thành K nếp gấp:

Dtrain = Df old 1 Df old 2 . . . Df già K

• Mỗi tập hợp nếp gấp Df old k đóng vai trò của một Deval k đánh giá. K tập huấn luyện tương ứng: •

Dtrain 1 = Dtrain / Dfold 1

• Dtrain 2 = Dtrain / Dfold 2
• . . .

## Ridge Regression - How to find $\lambda$?

- In ridge regression, $\lambda$ is a hyperparameter that we need to tune.
- We randomly split the training data:

$$\mathcal{D}_{train} = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)\}$$

into $K$ folds:

$$\mathcal{D}_{train} = \mathcal{D}_{fold-1} \cup \mathcal{D}_{fold-2} \cup \ldots \cup \mathcal{D}_{fold-K}$$

- Each fold set $\mathcal{D}_{fold-k}$ plays the role of an evaluation $\mathcal{D}_{eval-k}$. The corresponding $K$ training sets:
  - $\mathcal{D}_{train-1} = \mathcal{D}_{train} \setminus \mathcal{D}_{fold-1}$
  - $\mathcal{D}_{train-2} = \mathcal{D}_{train} \setminus \mathcal{D}_{fold-2}$
  - ...
  - $\mathcal{D}_{train-K} = \mathcal{D}_{train} \setminus \mathcal{D}_{fold-K}$





## Hồi quy sườn - Cách tìm λ?

- Trong hồi quy sườn núi, λ là một siêu tham số mà chúng ta cần điều chỉnh. •

Chúng tôi chia ngẫu nhiên dữ liệu huấn luyện:

Dtrain = {(x1, y1), . . . , (xn, yn)}

thành K nếp gấp:

Dtrain = Df old 1 Df old 2 . . . Df già K

- Mỗi tập hợp nếp gấp Df old k đóng vai trò của một Deval k đánh giá. K tập huấn luyện tương ứng: •

Dtrain 1 = Dtrain / Dfold 1

- Dtrain 2 = Dtrain / Dfold 2
- . . .
- Dtrain K = Dtrain / Dfold K

# Ridge Regression - How to find $\lambda$?

1. For $\lambda_b = 0.001, 0.002, \ldots, \lambda_B$:





# Hồi quy sườn - Cách tìm λ?

1. Với λb = 0,001, 0,002, . . . , λB:

VIET NAM NATIONAL UNIVERSITY HCMC
UNIVERSITY OF INFORMATION TECHNOLOGY
VNUHCM - UIT
17 / 19

# Ridge Regression - How to find $\lambda$?

1. For $\lambda_b = 0.001, 0.002, \ldots, \lambda_B$:
   - For $k = 1, \ldots, K$:





# Hồi quy sườn - Cách tìm λ?

1. Với λb = 0,001, 0,002, . . . , λB:
   - Với k = 1, . . . ,K:

# Ridge Regression - How to find $\lambda$?

1. For $\lambda_b = 0.001, 0.002, \ldots, \lambda_B$:
   - For $k = 1, \ldots, K$:
     - Fit RR model $h_{b,k}$ with $\lambda_b$ on $\mathcal{D}_{train-k}$





# Hồi quy sườn - Cách tìm λ?

1. Với λb = 0,001, 0,002, . . . , λB:
   - Với k = 1, . . . ,K:
     - Khớp mô hình RR hb,k với λb trên Dtrain k

# Ridge Regression - How to find $\lambda$?

1. For $\lambda_b = 0.001, 0.002, \ldots, \lambda_B$:
   - For $k = 1, \ldots, K$:
     - Fit RR model $h_{b,k}$ with $\lambda_b$ on $\mathcal{D}_{train-k}$
     - Compute and store $E_{eval-k}(h_{b,k})$ using $\mathcal{D}_{eval-k}$





# Hồi quy sườn - Cách tìm λ?

1. Với λb = 0,001, 0,002, . . . , λB:
   - Với k = 1, . . . ,K:
     - Khớp mô hình RR hb,k với λb trên Dtrain k
     - Tính toán và lưu trữ Eeval k(hb,k) bằng Deval k

# Ridge Regression - How to find $\lambda$?

1. For $\lambda_b = 0.001, 0.002, \ldots, \lambda_B$:
   - For $k = 1, \ldots, K$:
     - Fit RR model $h_{b,k}$ with $\lambda_b$ on $\mathcal{D}_{train-k}$
     - Compute and store $E_{eval-k}(h_{b,k})$ using $\mathcal{D}_{eval-k}$
   - Compute and store $E_{cv_b} = \frac{1}{K} \sum_k E_{eval-k}(h_{b,k})$





# Hồi quy sườn - Cách tìm λ?

1. Với λb = 0,001, 0,002, . . . , λB:
   - Với k = 1, . . . ,K:
     - Khớp mô hình RR hb,k với λb trên Dtrain k
     - Tính toán và lưu trữ Eeval k(hb,k) bằng Deval k
   - Tính toán và lưu trữ Ecvb = $\frac{1}{K}$ k Eeval k(hb,k)

# Ridge Regression - How to find $\lambda$?

1. For $\lambda_b = 0.001, 0.002, \ldots, \lambda_B$:
   - For $k = 1, \ldots, K$:
     - Fit RR model $h_{b,k}$ with $\lambda_b$ on $\mathcal{D}_{train-k}$
     - Compute and store $E_{eval-k}(h_{b,k})$ using $\mathcal{D}_{eval-k}$
   - Compute and store $E_{cv_b} = \frac{1}{K}\sum_k E_{eval-k}(h_{b,k})$
2. Compute all cross validation errors $E_{cv_1}, E_{cv_2}, \ldots, E_{cv_B}$ and choose the smallest $E_{cv_{b^*}}$.





# Hồi quy sườn - Cách tìm λ?

1. Với λb = 0,001, 0,002, . . . , λB: •
   Với k = 1, . . . ,K:
   • Khớp mô hình RR hb,k với λb trên
   Dtrain k • Tính toán và lưu trữ Eeval k(hb,k)
   bằng Deval k • Tính toán và lưu trữ Ecvb = $\frac{1}{K}$ k Eeval k(hb,k)
2. Tính toán tất cả các lỗi xác thực chéo Ecv1 , Ecv2 , . . . , EcvB và chọn Ecvb nhỏ nhất .

# Ridge Regression - How to find $\lambda$?

1. For $\lambda_b = 0.001, 0.002, \ldots, \lambda_B$:
   - For $k = 1, \ldots, K$:
     - Fit RR model $h_{b,k}$ with $\lambda_b$ on $\mathcal{D}_{train-k}$
     - Compute and store $E_{eval-k}(h_{b,k})$ using $\mathcal{D}_{eval-k}$
   - Compute and store $E_{cv_b} = \frac{1}{K} \sum_k E_{eval-k}(h_{b,k})$
2. Compute all cross validation errors $E_{cv_1}, E_{cv_2}, \ldots, E_{cv_B}$ and choose the smallest $E_{cv_{b^*}}$.
3. Use $\lambda^*$ to fit the final Ridge Regression model:

$$\hat{\mathbf{y}} = (\mathbf{X}^\top \mathbf{X} + \lambda^* \mathbf{I})^{-1} \mathbf{X}^\top \mathbf{y}$$





# Hồi quy sườn - Cách tìm λ?

1 Với λb = 0,001, 0,002, . . . , λB: •

Với k = 1, . . . ,K:

• Khớp mô hình RR hb,k với λb trên

Dtrain k • Tính toán và lưu trữ Eeval k(hb,k)

bằng Deval k • Tính toán và lưu trữ Ecvb = 1/K k Eeval k(hb,k)

2 Tính toán tất cả các lỗi xác thực chéo Ecv1 , Ecv2 , . . . , EcvB và chọn Ecvb nhỏ nhất .

3 Sử dụng λ để phù hợp với mô hình Hồi quy sườn cuối cùng:

y^ = (X X + λ tôi) 1X y

# Least Absolute Shrinkage and Selection Operator - LASSO

- Instead of using $L_2$ norm as in Ridge Regression:

$$\min_{\mathbf{b}\in\mathbb{R}^p}\left\{\frac{1}{n}\|\mathbf{X}\mathbf{b}-\mathbf{y}\|_2^2\right\} \quad \texttt{subject to} \quad \|\mathbf{b}\|_2^2 \le c$$





# Toán tử lựa chọn và co rút tuyệt đối nhỏ nhất - LASSO

- Thay vì sử dụng định mức L2 như trong Hồi quy độ dốc:

tối thiểu b Rp {1/N Xb y 2 2} theo b 2 ≤ c 2

# Least Absolute Shrinkage and Selection Operator - LASSO

- Instead of using $L_2$ norm as in Ridge Regression:

$$\min_{\mathbf{b}\in\mathbb{R}^p}\left\{\frac{1}{n}\|\mathbf{X}\mathbf{b}-\mathbf{y}\|_2^2\right\} \quad \texttt{subject to} \quad \|\mathbf{b}\|_2^2 \le c$$

- We use $L_1$ norm:

$$\min_{\mathbf{b}\in\mathbb{R}^p}\left\{\frac{1}{n}\|\mathbf{X}\mathbf{b}-\mathbf{y}\|_2^2\right\} \quad \texttt{subject to} \quad \|\mathbf{b}\|_1 \le c$$





# Toán tử lựa chọn và co rút tuyệt đối nhỏ nhất - LASSO

- Thay vì sử dụng định mức L2 như trong Hồi quy độ dốc:

tối thiểu b Rp {1/N Xb y 2 2} theo b 2 ≤ c 2

- Ta sử dụng định mức L1 :

tối thiểu b Rp {1/N Xb y 2 2} theo b 1 ≤ c

# Least Absolute Shrinkage and Selection Operator - LASSO

- Instead of using $L_2$ norm as in Ridge Regression:

$$\min_{\mathbf{b}\in\mathbb{R}^p} \left\{ \frac{1}{n} \|\mathbf{X}\mathbf{b} - \mathbf{y}\|_2^2 \right\} \quad \texttt{subject to} \quad \|\mathbf{b}\|_2^2 \leq c$$

- We use $L_1$ norm:

$$\min_{\mathbf{b}\in\mathbb{R}^p} \left\{ \frac{1}{n} \|\mathbf{X}\mathbf{b} - \mathbf{y}\|_2^2 \right\} \quad \texttt{subject to} \quad \|\mathbf{b}\|_1 \leq c$$

- The $L_1$ norm constraint is:

$$\|\mathbf{b}\|_1 \leq c \iff \sum_{j=1}^{p} |b_j| \leq c$$





# Toán tử lựa chọn và co rút tuyệt đối nhỏ nhất - LASSO

- Thay vì sử dụng định mức L2 như trong Hồi quy độ dốc:

tối thiểu 1
b Rp {N Xb y 2 2} theo b 2 ≤ c 2

- Ta sử dụng định mức L1 :

tối thiểu 1
b Rp {N Xb y 2 2} theo b 1 ≤ c

- Ràng buộc định mức L1 là:

b 1 ≤ c P bj ≤ c
j=1

# Least Absolute Shrinkage and Selection Operator - LASSO

- Instead of using $L_2$ norm as in Ridge Regression:

$$\min_{\mathbf{b}\in\mathbb{R}^p}\left\{\frac{1}{n}\|\mathbf{Xb}-\mathbf{y}\|_2^2\right\} \quad \texttt{subject to} \quad \|\mathbf{b}\|_2^2 \le c$$

- We use $L_1$ norm:

$$\min_{\mathbf{b}\in\mathbb{R}^p}\left\{\frac{1}{n}\|\mathbf{Xb}-\mathbf{y}\|_2^2\right\} \quad \texttt{subject to} \quad \|\mathbf{b}\|_1 \le c$$

- The $L_1$ norm constraint is:

$$\|\mathbf{b}\|_1 \le c \iff \sum_{j=1}^{p} |b_j| \le c$$

- Our minimization problem then becomes:

$$\min_{\mathbf{b}\in\mathbb{R}^p}\left\{\frac{1}{n}(\mathbf{Xb}-\mathbf{y})^\top(\mathbf{Xb}-\mathbf{y}) + \lambda\sum_{j=1}^{p}|b_j|\right\}$$





# Toán tử lựa chọn và co rút tuyệt đối nhỏ nhất - LASSO

- Thay vì sử dụng định mức L2 như trong Hồi quy độ dốc:

tối thiểu b Rp {1/N Xb y 2 2} theo b 2 ≤ c 2

- Ta sử dụng định mức L1 :

tối thiểu b Rp {1/N Xb y 2 2} theo b 1 ≤ c

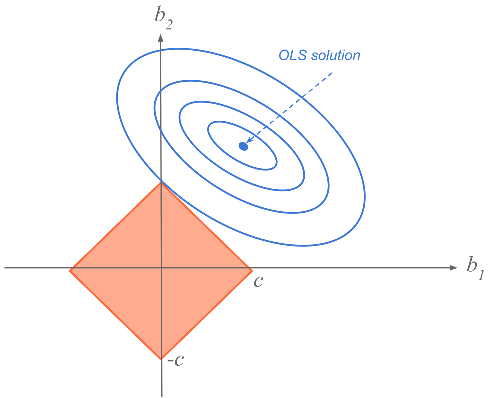
- Ràng buộc định mức L1 là:

b 1 ≤ c P bj ≤ c j=1

- Khi đó bài toán tối thiểu hóa của chúng ta trở thành:

tối thiểu b Rp 1/N (Xb y) (Xb y) + λ P bj j=1

# LASSO - Variable Selection

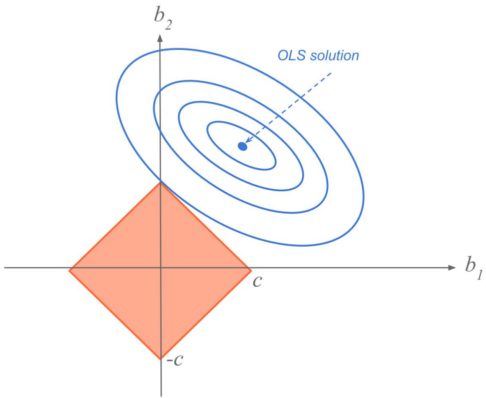



- The ideal point has $b_1$ coordinate equal to 0.
- LASSO completely zero-ed out $b_1$ in the model, reducing the number of coefficients from 2 down to 1.





# LASSO - Lựa chọn biến




• Điểm lý tưởng có tọa độ b1 bằng 0. • LASSO loại bỏ hoàn

toàn b1 trong mô hình, giảm
số hệ số từ 2 xuống còn 1.