

دوره
علم داده،
یادگیری ماشین و
هوش مصنوعی

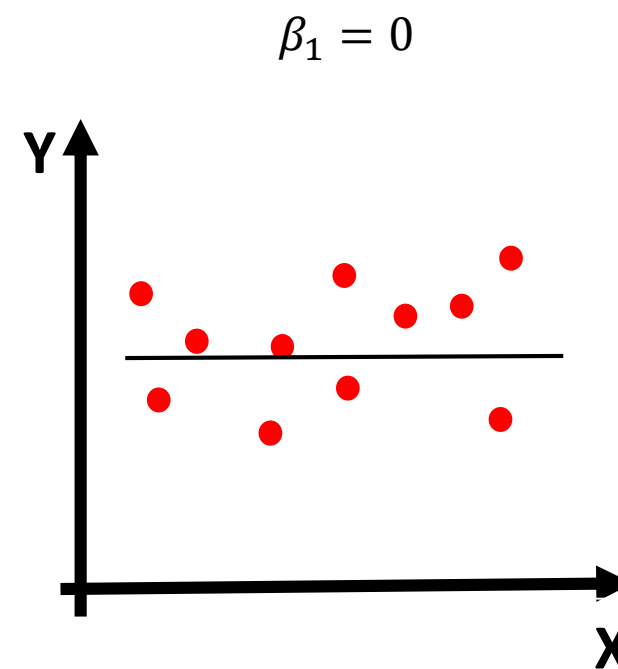
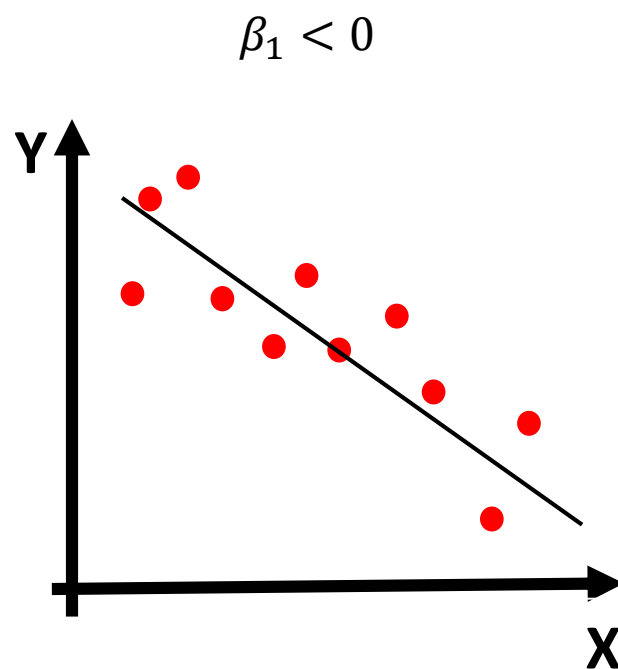
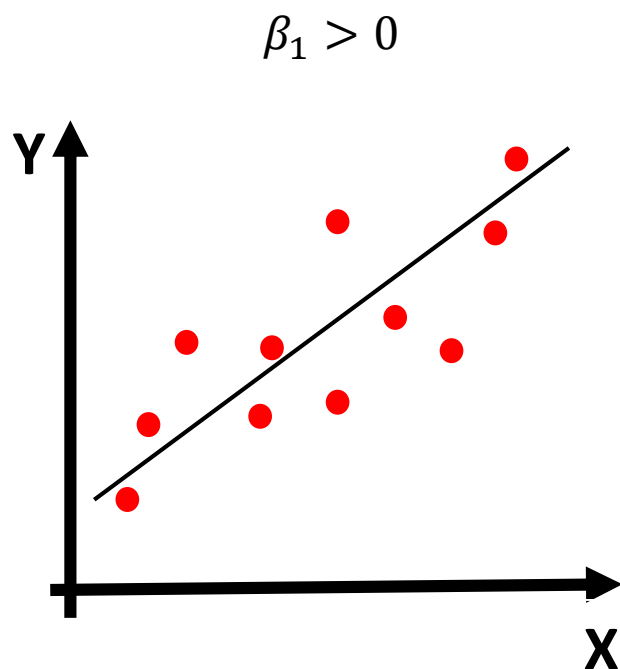
جلسه یازدهم:
الگوریتم رگرسیون خطی

ارائه‌کننده:
دکتر فرزاد مینویی

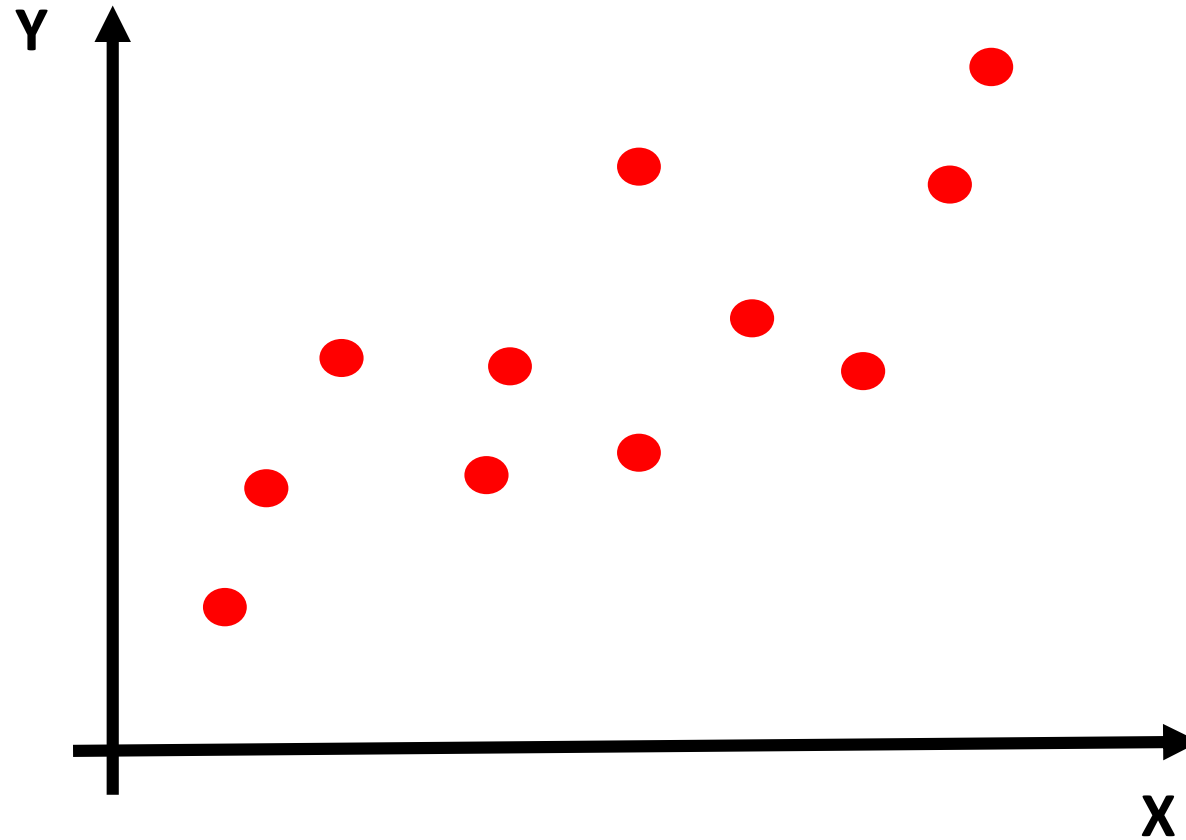


رگرسیون خطی ساده

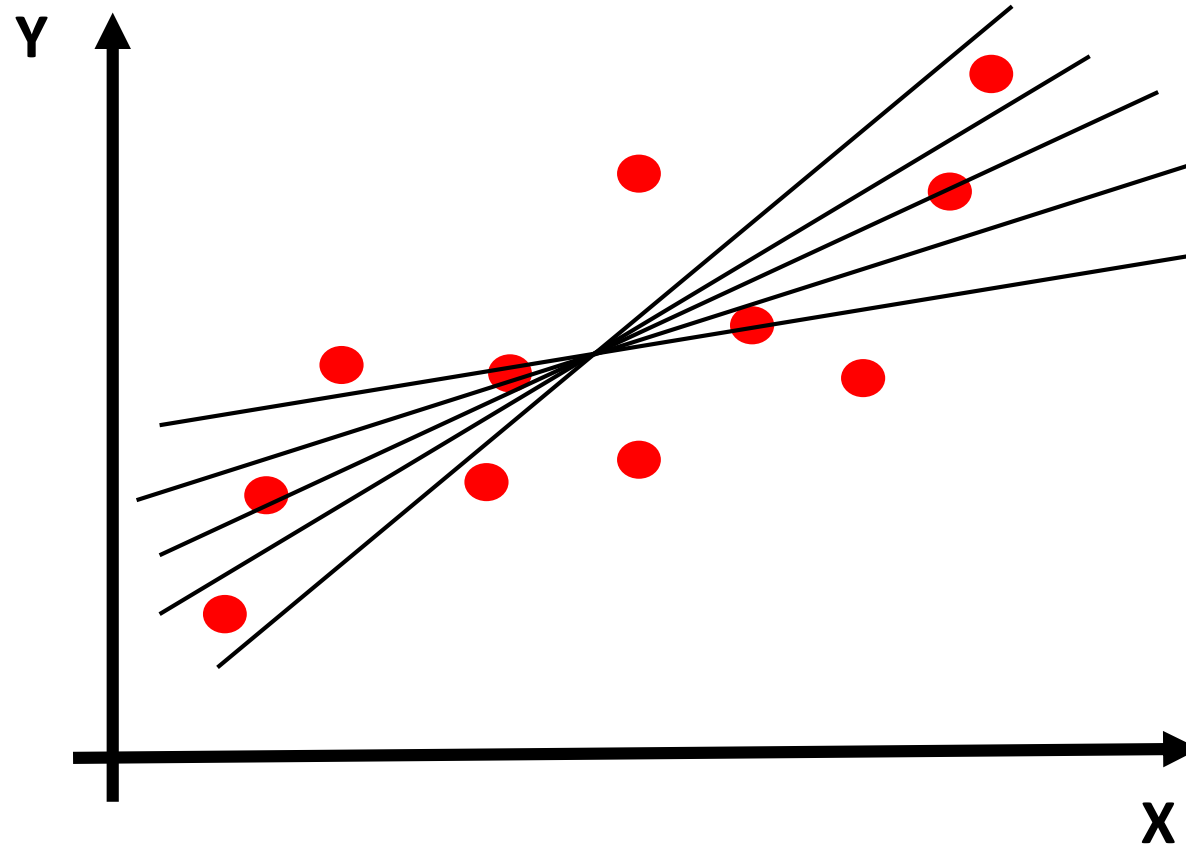
رابطه واقعی بین x و y در جامعه: $y = \beta_0 + \beta_1 x + \varepsilon$



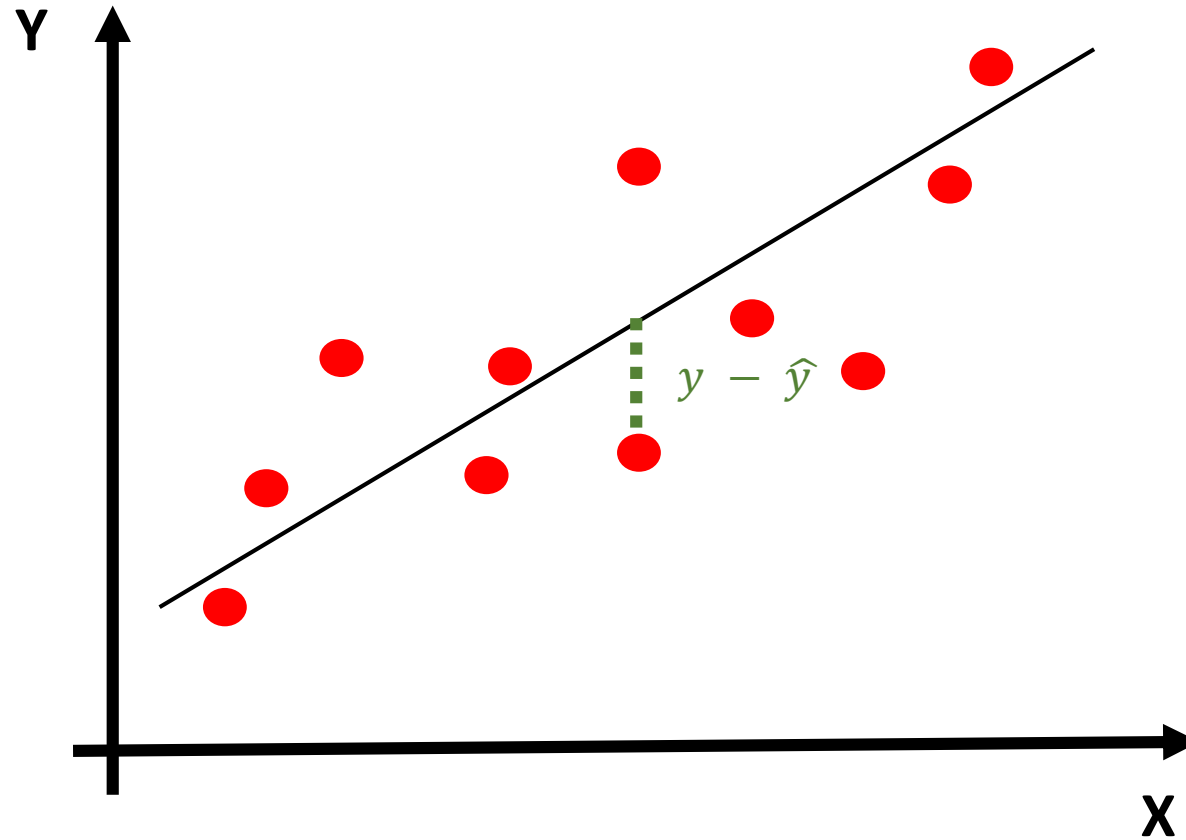
برازش خط از بین نقاط



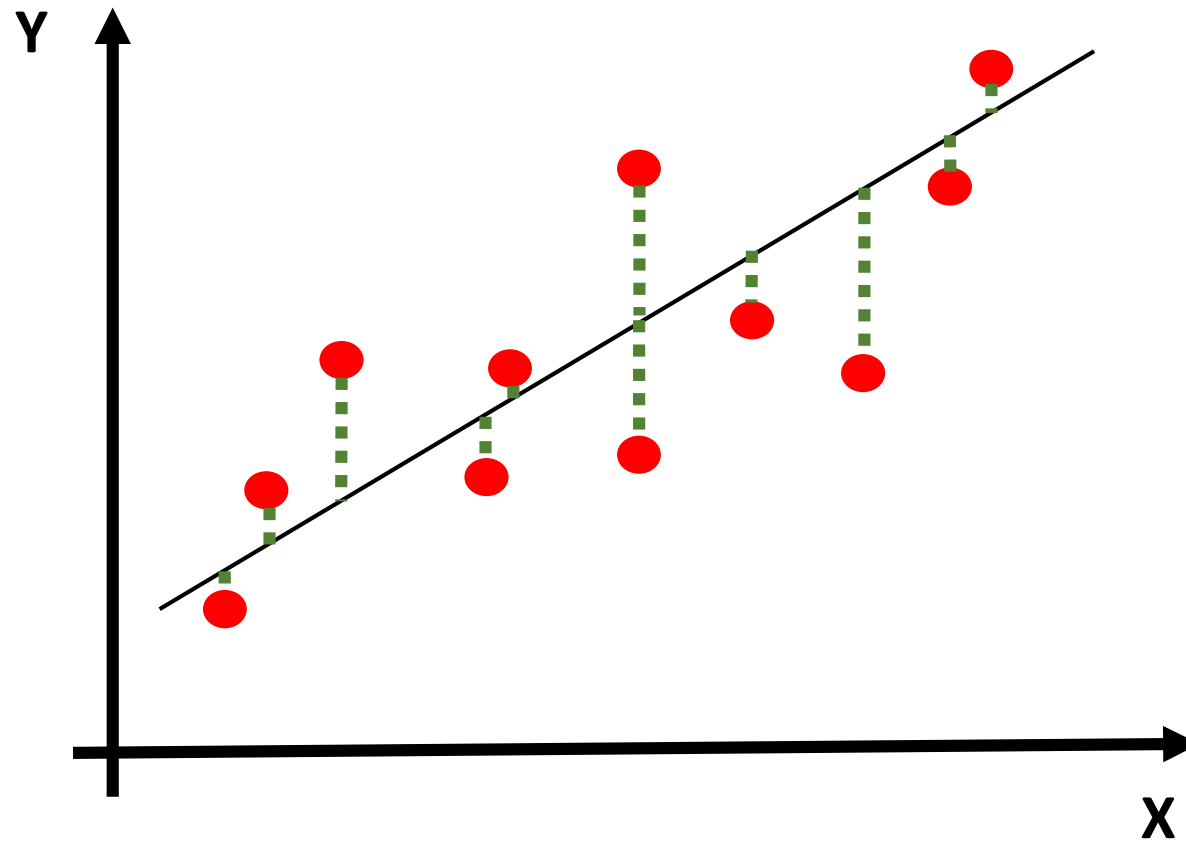
برازش خط از بین نقاط



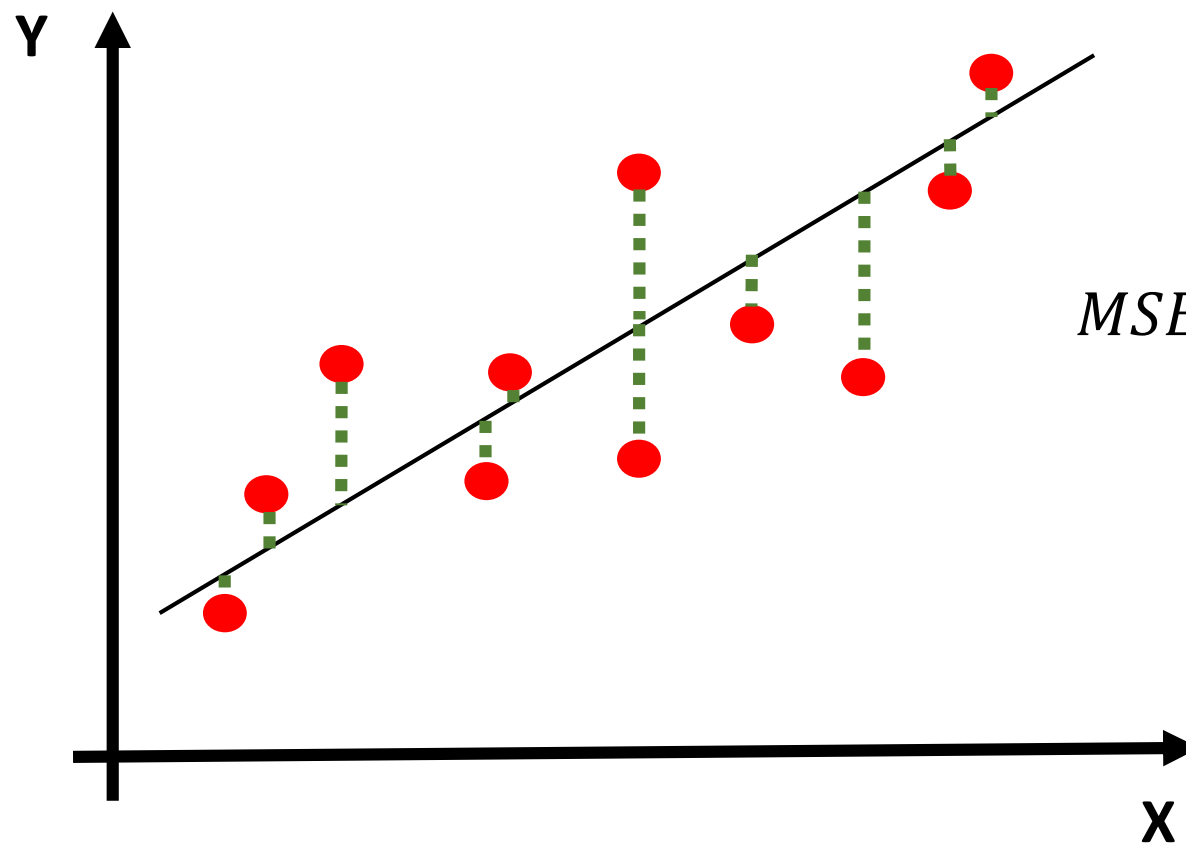
روش حداقل مجذورات (Ordinary Least Squares - OLS)



روش حداقل مجزورات



روش حداقل مجزورات



$$MSE = \frac{1}{n} \sum (y_i - \hat{y}_i)^2$$

هوش محاسباتی مجذورات

تابع هزینه (Cost Function)

$$\text{Minimize}_{\beta}: MSE = \frac{1}{n} \sum (y_i - \hat{y}_i)^2 = \frac{1}{n} \sum (y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))^2$$

صورت‌بندی ماتریسی رگرسیون خطی برای آموزش

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + \hat{\varepsilon}_i$$

$$Y = X \hat{\beta}$$

داده آموزش

| y | x |
|----|----|
| 80 | 8 |
| 74 | 12 |
| 61 | 14 |
| 47 | 10 |
| 91 | 23 |



ماتریس ویژگی‌ها

$$\begin{bmatrix} 1 & 8 \\ 1 & 12 \\ 1 & 14 \\ 1 & 10 \\ 1 & 23 \end{bmatrix}$$

\times

ماتریس ضرائب

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix}$$

$=$

ماتریس پاسخ

$$\begin{bmatrix} 80 \\ 74 \\ 61 \\ 47 \\ 91 \end{bmatrix}$$

حل رگرسیون خطی با رویکرد محاسبه دقیق

• اگر ماتریس X مربعی و وارون پذیر باشد

$$Y = X \hat{\beta}$$

$$X^{-1} Y = X^{-1} X \hat{\beta} \rightarrow X^{-1} Y = I \hat{\beta} \rightarrow \hat{\beta} = X^{-1} Y$$

• رویکرد Normal Equation

$$Y = X \hat{\beta}$$

$$X^T Y = X^T X \hat{\beta} \rightarrow (X^T X)^{-1} X^T Y = (X^T X)^{-1} (X^T X) \hat{\beta} \rightarrow$$

$$(X^T X)^{-1} X^T Y = I \hat{\beta} \rightarrow \hat{\beta} = (X^T X)^{-1} X^T Y$$

حل رگرسیون خطی با رویکرد محاسبه دقیق

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

• رویکرد Normal Equation

• در شرایطی که ماتریس $X^T X$ وارون پذیر نباشد، کار نمی‌کند.

• هم‌خطی کامل بین متغیرهای پیش‌بینی کننده (Multicollinearity)

• در حالتی که n (sample size) $< m$ (# of features)

$$X = U \Sigma V^T \rightarrow \hat{\beta} = V \Sigma^{-1} U^T Y$$

• رویکرد Singular Value Decomposition (SVD)

• توسط کلاس `sklearn.linear_model.LinearRegression` پشتیبانی می‌شود.

• زمان محاسبات برای آموزش مدل $O(m^2 \times n_{sample} + m^3)$

• قاعده سرانگشتی برای حداقل نمونه مورد نیاز برای آموزش رگرسیون خطی $30 \times m$ (# of features)

صورت‌بندی ماتریسی رگرسیون خطی برای پیش‌بینی

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

$$\hat{y}_i = 49.31 + 1.59 x_i$$

$$\hat{Y} = X \hat{\beta}$$

داده جدید

| |
|----|
| x |
| 5 |
| 8 |
| 21 |
| 11 |
| 14 |



ماتریس ویژگی‌ها

$$\begin{bmatrix} 1 & 5 \\ 1 & 8 \\ 1 & 21 \\ 1 & 11 \\ 1 & 14 \end{bmatrix}$$

×

ماتریس ضرائب

$$\begin{bmatrix} 49.31 \\ 1.59 \end{bmatrix}$$

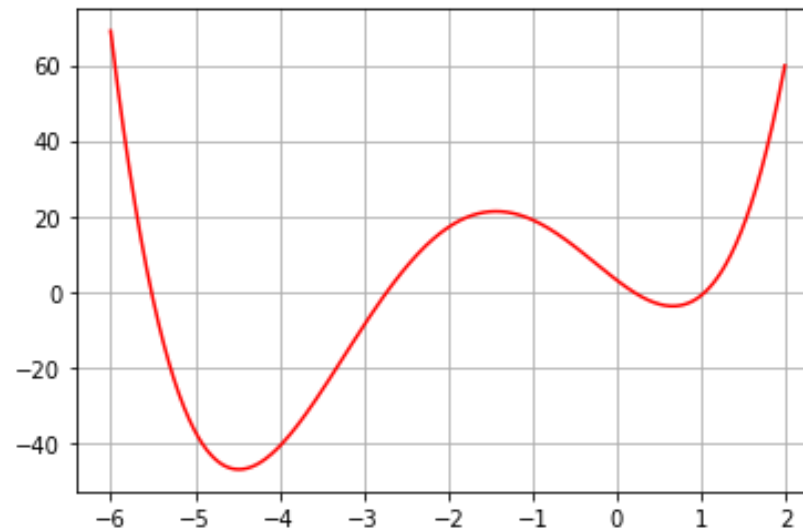
=

ماتریس پاسخ

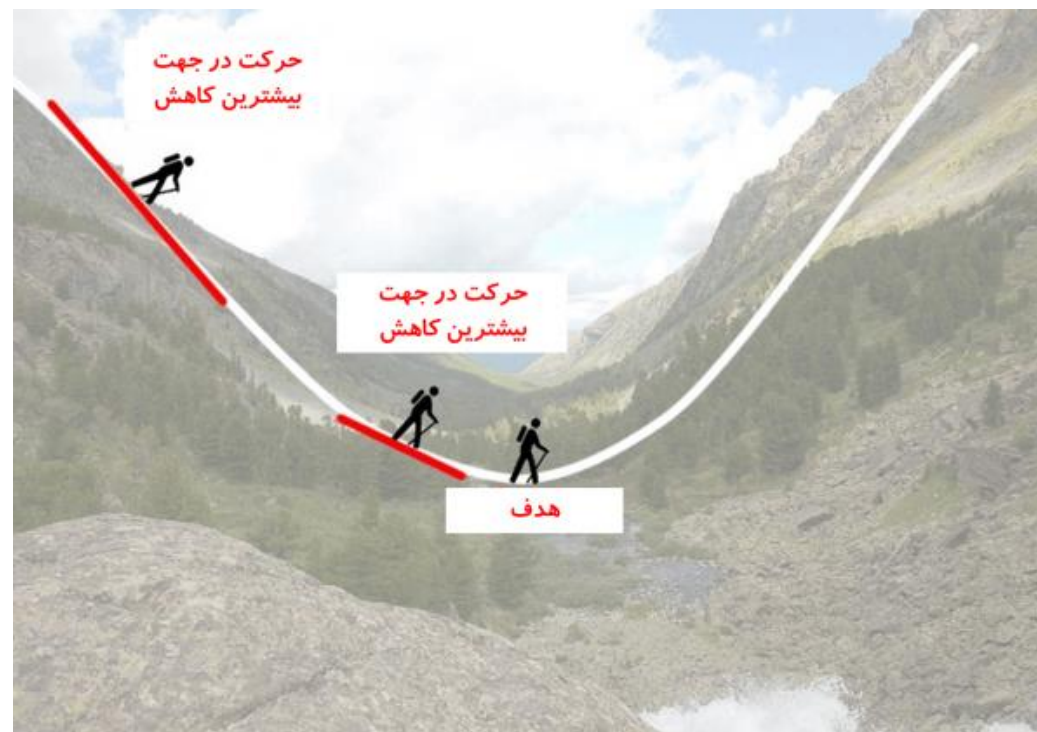
$$\begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \hat{y}_3 \\ \hat{y}_4 \\ \hat{y}_5 \end{bmatrix}$$

رویکردهای عددی برای حل مسائل بهینه‌سازی

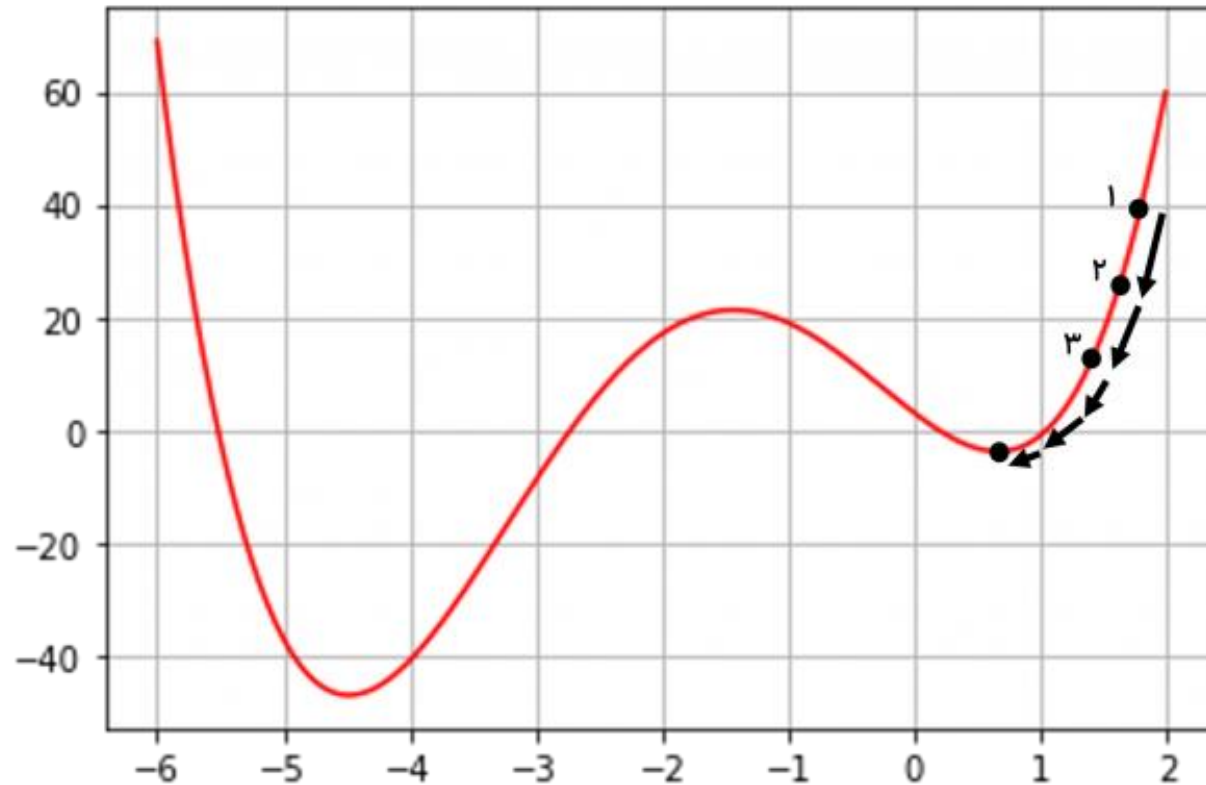
• الگوریتم گرادیان کاهشی (Gradient Descent)



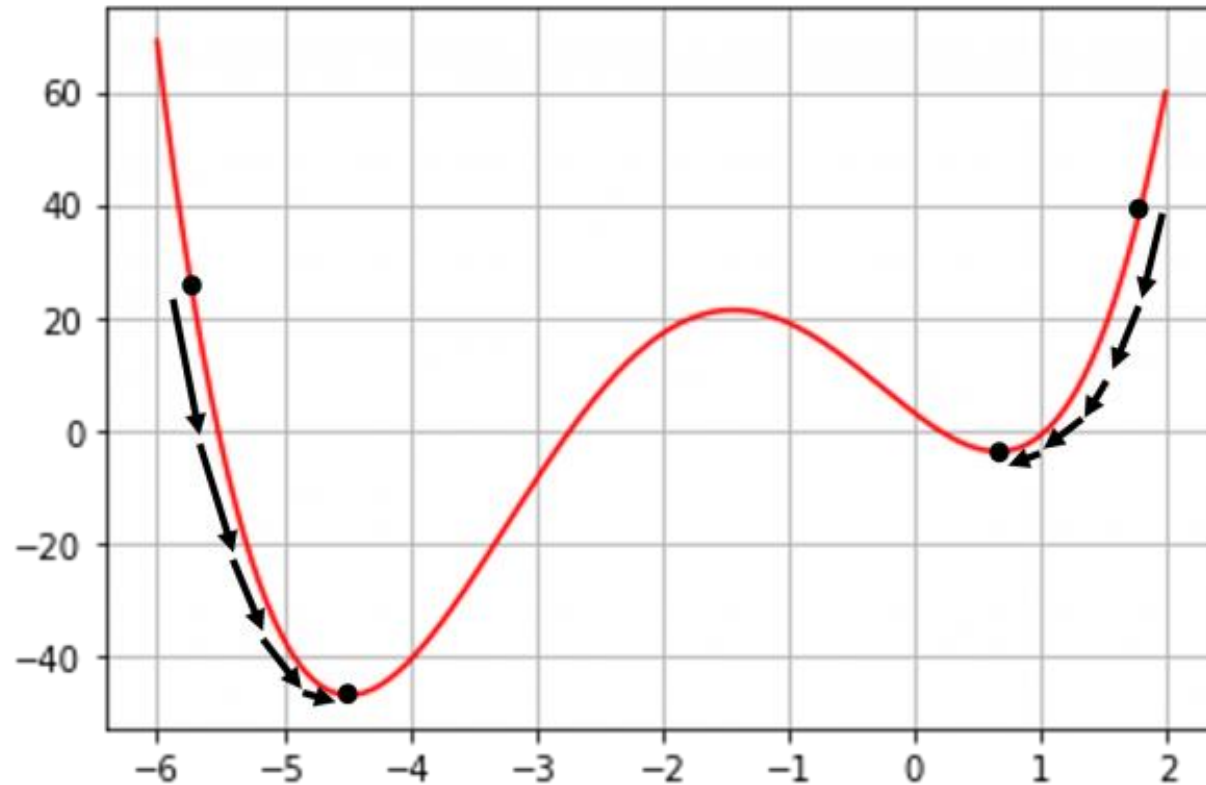
فهم شهودی الگوریتم گرادیان کاهش



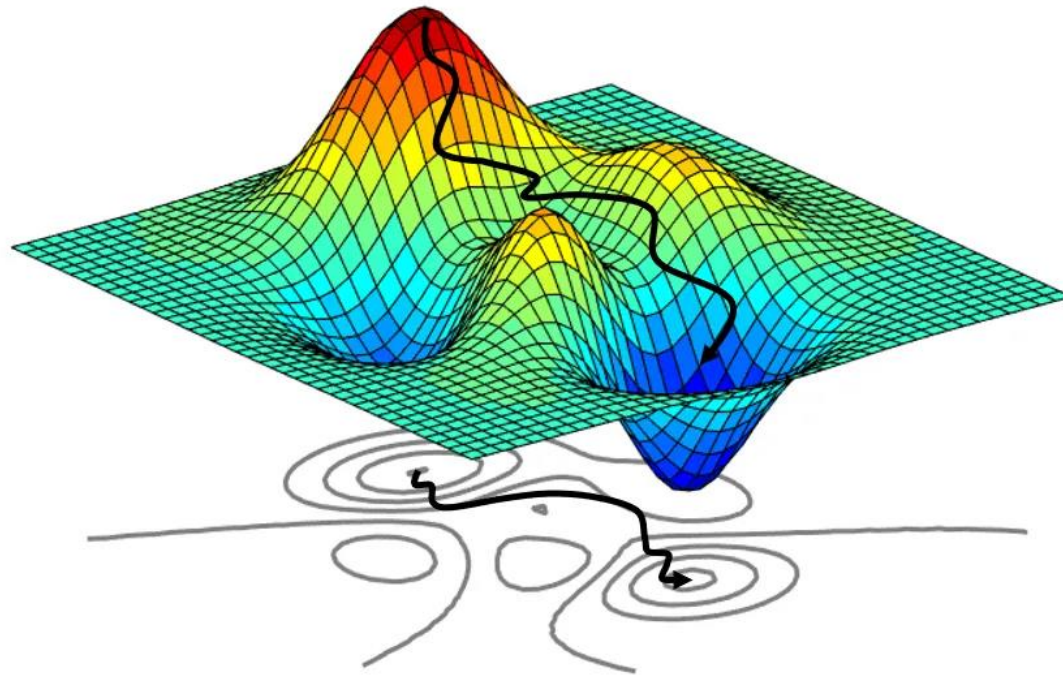
فهم شهودی الگوریتم گرادیان کاهشی



فهم شهودی الگوریتم گرادیان کاهشی

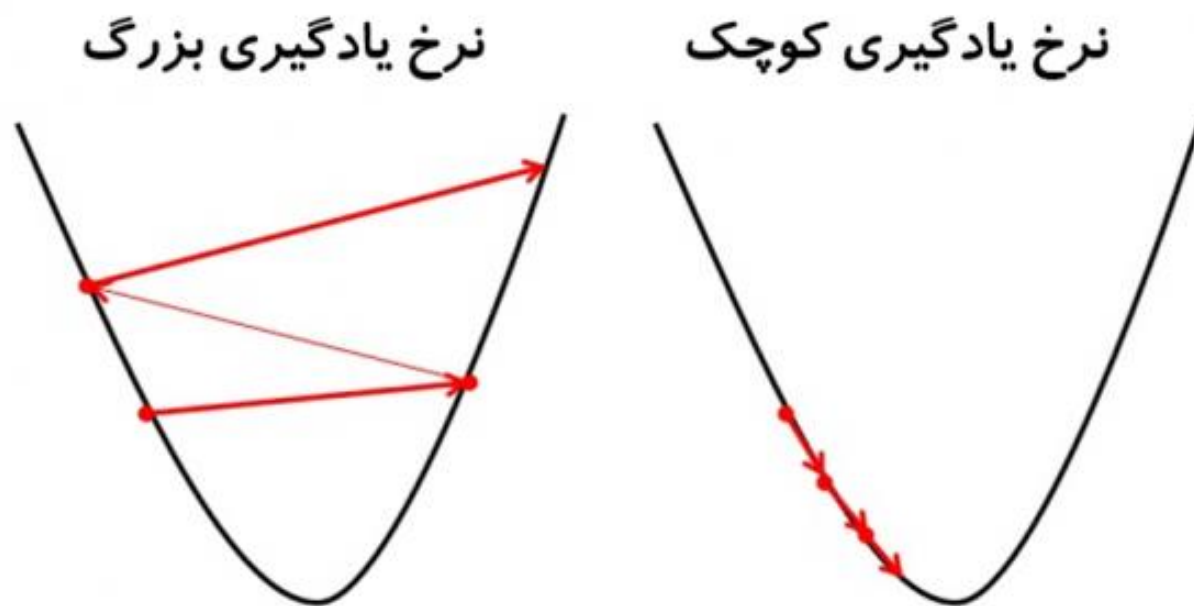


الگوریتم گرادیان کاهشی



$$X_{i+1} = X_i - \text{learning rate} \times \nabla f(X_i)$$

نرخ یادگیری (Learning Rate)



حل رگرسیون خطی با رویکرد گرادیان کاهشی

$$\hat{\beta}^{i+1} = \hat{\beta}^i - \text{learning rate} \times \nabla \text{MSE}(\hat{\beta}^i)$$

• رویکرد گرادیان کاهشی

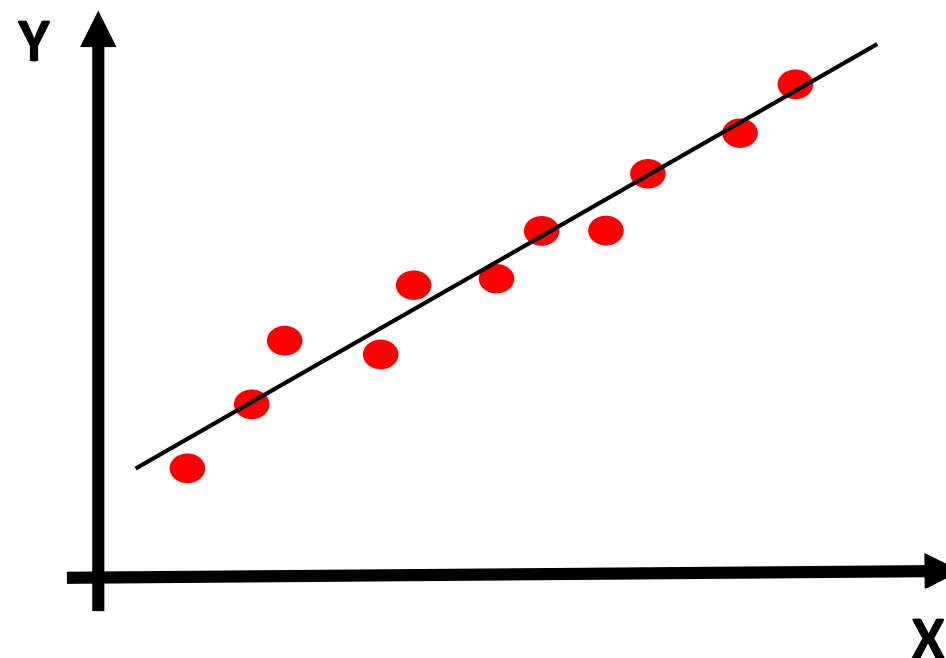
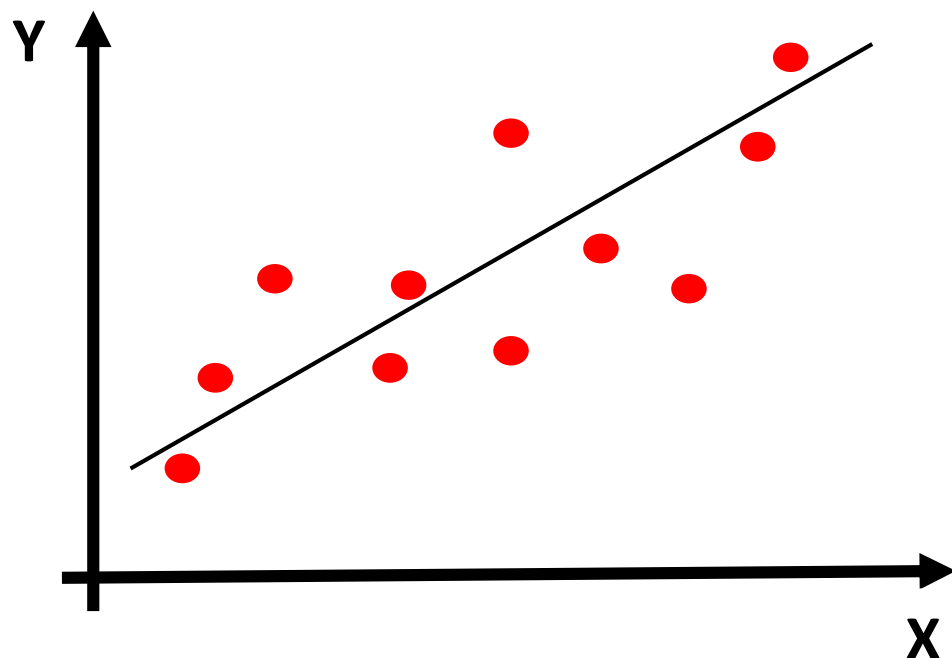
• تابع هزینه در رگرسیون خطی یک تابع محدب (Convex Function) است، بنابراین تنها یک کمینه جهانی (Global Minimum) دارد.

• کلاس `sklearn.linear_model.SGDRegressor` از رویکرد Stochastic Gradient Descent استفاده می‌کند.

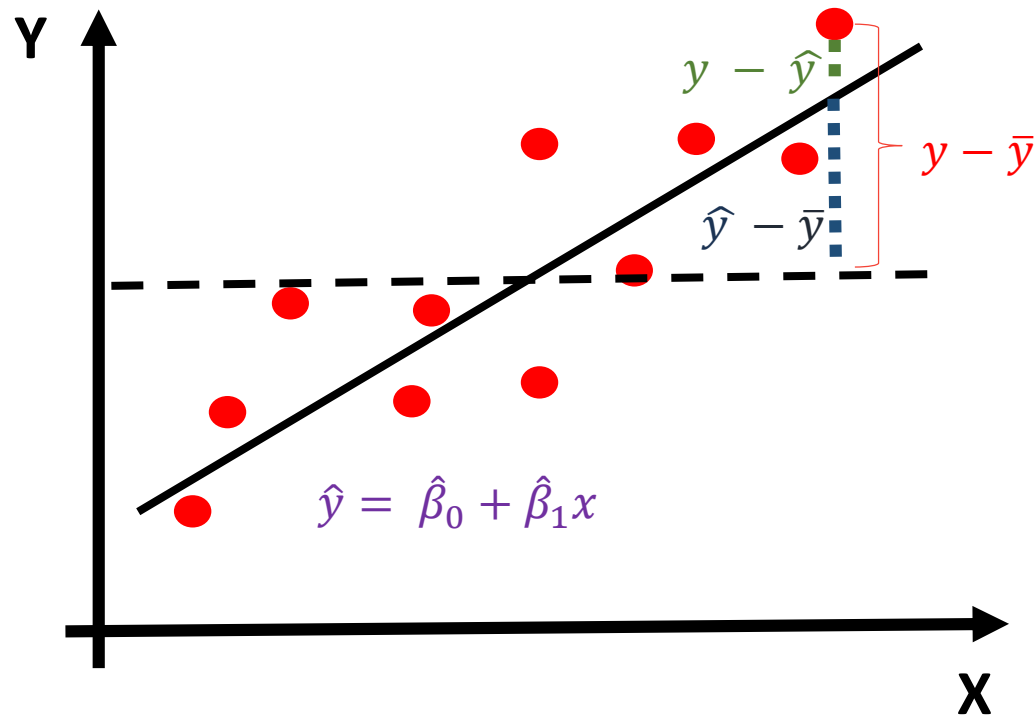
• برای اطمینان از همگرایی الگوریتم، نیاز است همه ویژگی‌ها مقیاس شوند.

• نرخ یادگیری معمولاً بین ۰.۰۰۱ تا ۰.۳ در نظر گرفته می‌شود.

چقدر مدل خوب برازش شده است؟



چقدر مدل خوب برازش شده است؟



SSE (Sum of Squares due to Error)

$$SSE = \sum (y_i - \hat{y}_i)^2$$

SST (Total Sum of Squares)

$$SST = \sum (y_i - \bar{y})^2$$

SSR (Sum of Squares due to Regression)

$$SSR = \sum (\hat{y}_i - \bar{y})^2$$

$$SST = SSR + SSE$$

Coefficient of Determination

$$R^2 = \frac{SSR}{SST}, 0 \leq R^2 \leq 1$$

رگرسیون خطی چندمتغیره (Multiple Linear Regression)

- وقتی تعداد متغیرهای پیش‌بینی‌کننده بیش از یکی باشد، چه کنیم؟

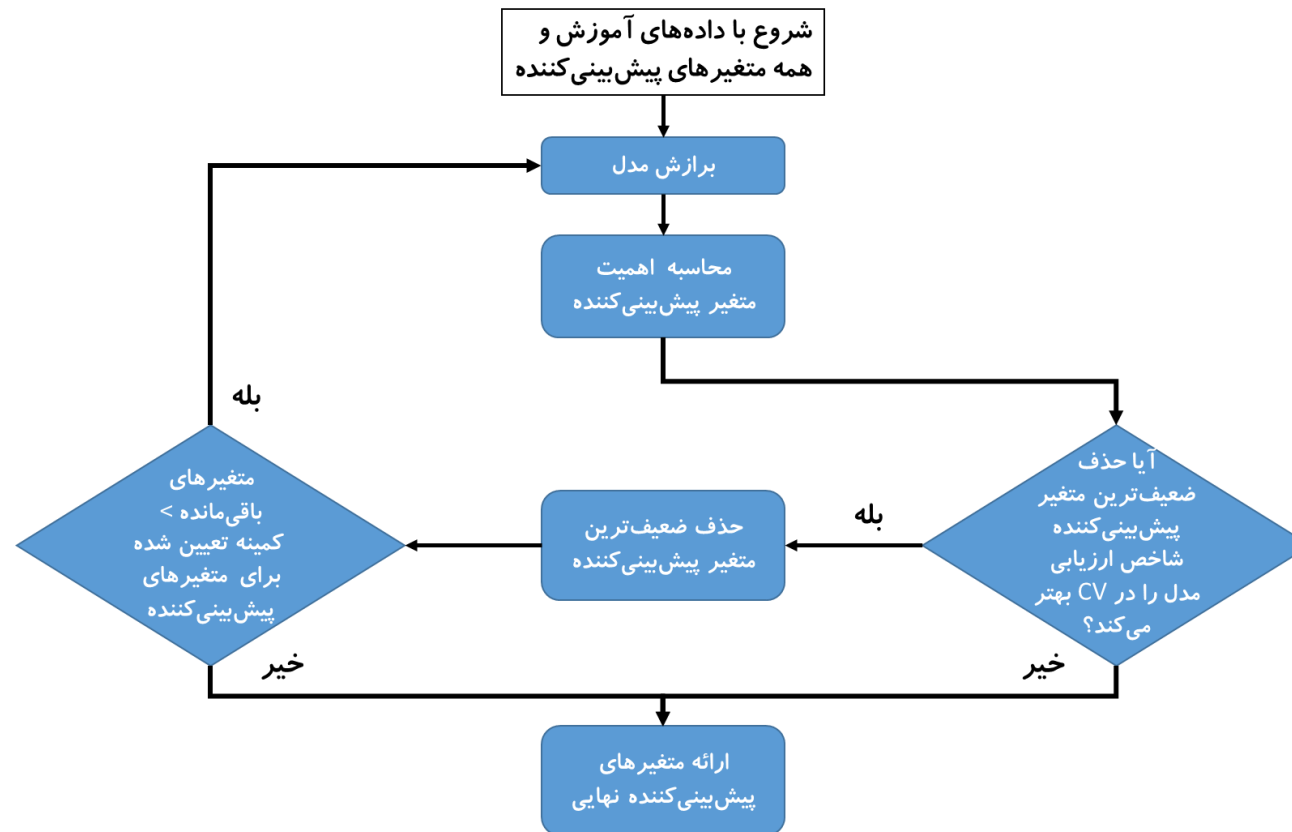
- چرا به ازای هر متغیر پیش‌بینی‌کننده یک مدل نمی‌سازیم؟

- رگرسیون خطی چندمتغیره

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_m x_m$$

انتخاب ویژگی‌ها (Feature Selection)

• الگوریتم Recursive Feature Elimination with Cross-Validation (RFECV)



متغیرهای رسته‌ای به عنوان پیش‌بینی‌کننده

- مثال پیش‌بینی سطح درآمد براساس سابقه کار (x_1) و جنسیت (x_2)
- تعریف متغیر مصنوعی (Dummy Variable) برای جنسیت
- حالت پایه: زن بودن

$$x_2 = \begin{cases} 0 & \text{if the person is female} \\ 1 & \text{if the person is male} \end{cases}$$

- رابطه رگرسیون خطی

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$$

متغیرهای رسته‌ای به عنوان پیش‌بینی‌کننده

- مثال پیش‌بینی سطح درآمد براساس سابقه کار (x_1) و سطح تحصیلات (x_2)
- سه حالت برای سطح تحصیلات: دبیرستان (L1)، کارشناسی و کارشناسی ارشد (L2) و دکترا (L3)
- تعریف دو متغیر مصنوعی برای سطح تحصیلات
- حالت پایه: L2

$$x_{21} = \begin{cases} 1 & \text{for L1} \\ 0 & \text{else.} \end{cases} \quad x_{22} = \begin{cases} 1 & \text{for L3} \\ 0 & \text{else.} \end{cases}$$

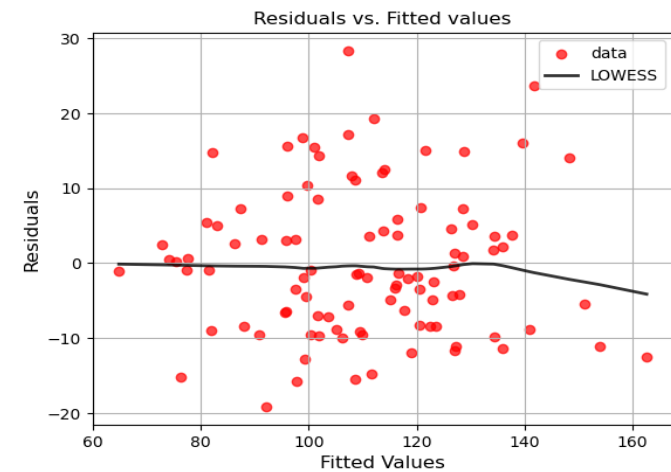
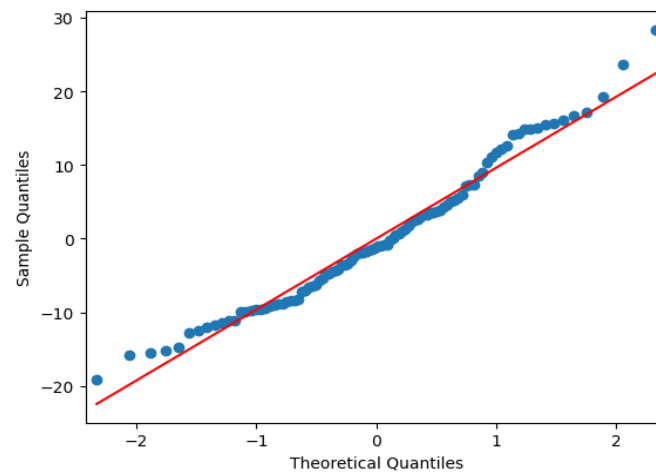
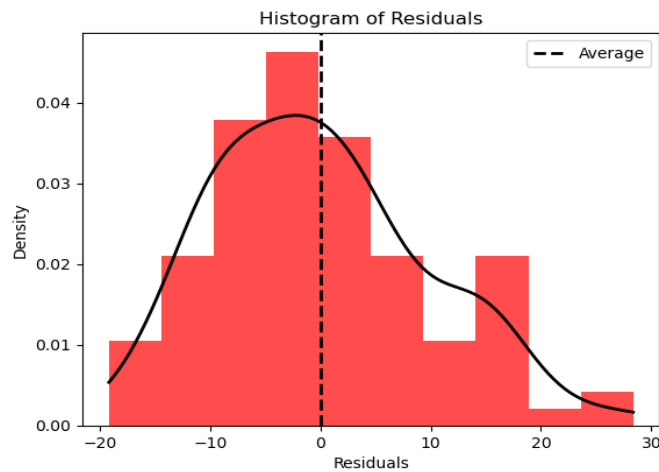
- رابطه رگرسیون خطی

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_{21} x_{21} + \hat{\beta}_{22} x_{22}$$

فرض الگوریتم رگرسیون خطی درباره خطاها

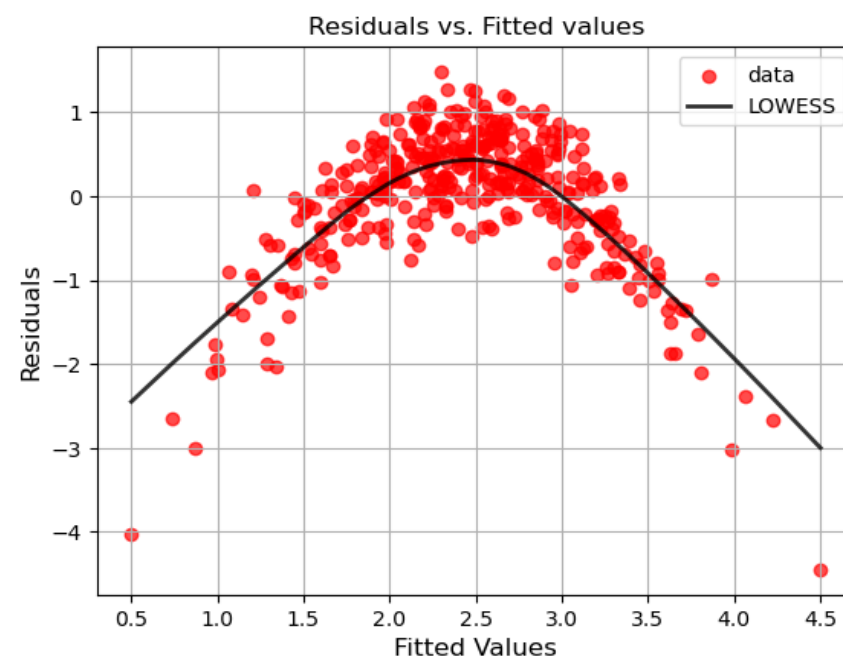
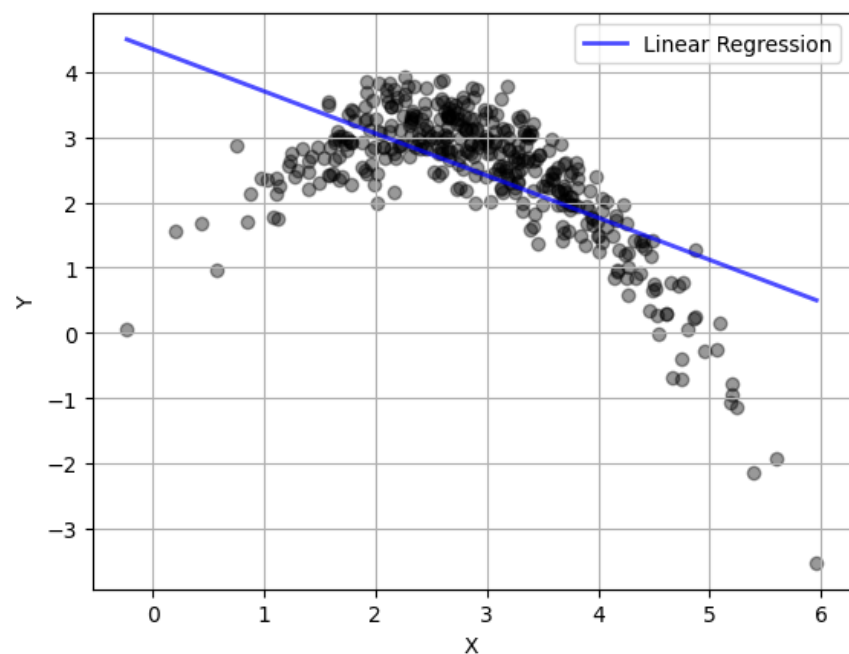
- مقدار خطا (ε) دارای توزیع نرمال با میانگین صفر و انحراف معیار ثابت است.

$$\varepsilon \sim N(0, \sigma^2)$$



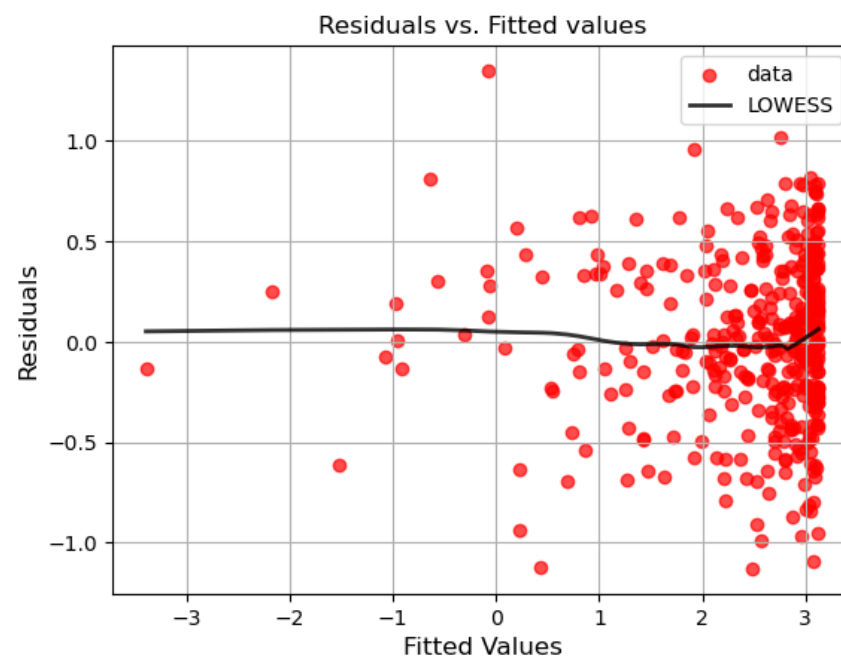
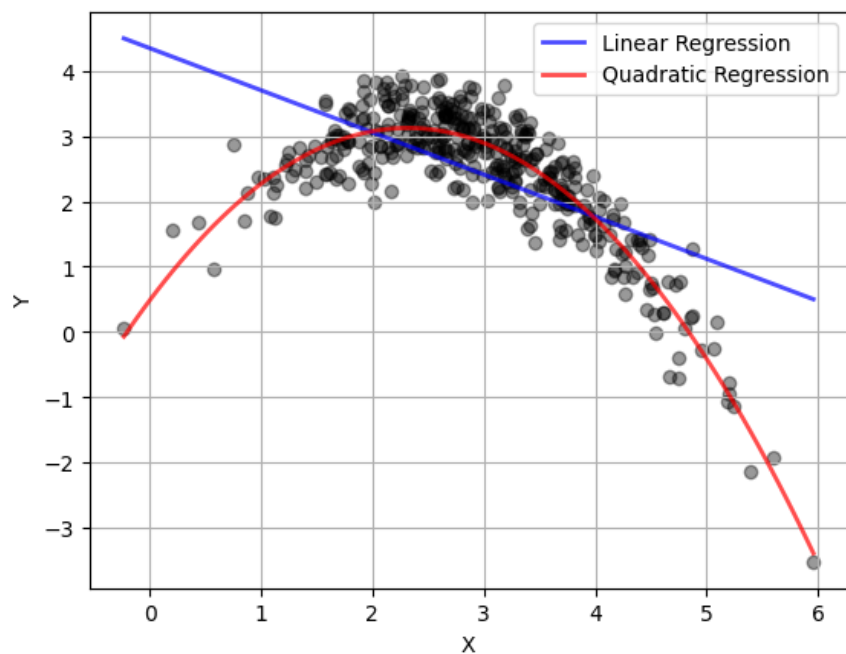
نقض فرض خطی بودن

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$



مدل سازی رابطه غیرخطی

• رگرسیون چندجمله‌ای (Polynomial Regression) $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1x + \hat{\beta}_2x^2$



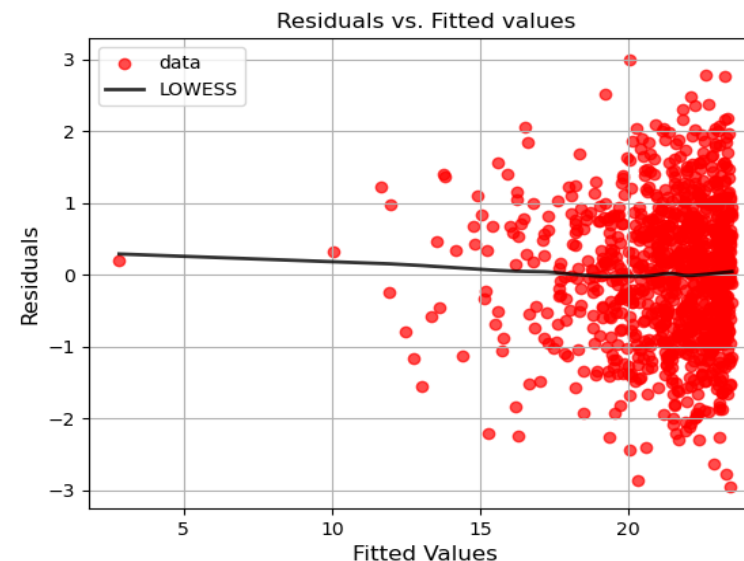
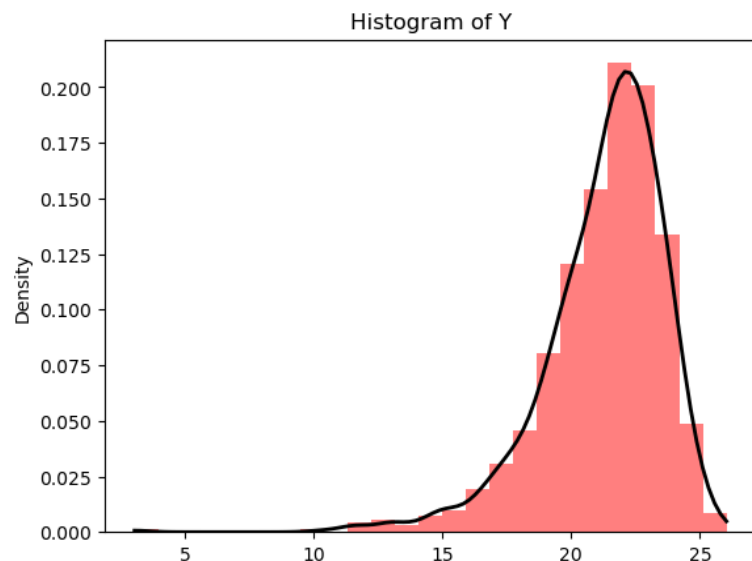
مدل سازی اثر متقابل (Interaction Effect)

- مفهوم اثر متقابل:
- اثر سیگار کشیدن بر روی ریسک ابتلاء به سرطان
- اثر در معرض فیبرهای آزبست بودن بر روی ریسک ابتلاء به سرطان
- اثر متقابل: اثر هم زمان سیگار کشیدن و در معرض آزبست بودن بیشتر از جمع دو اثر بالاست.

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_{12} x_1 \times x_2$$

تغییر متغیر Box-Cox

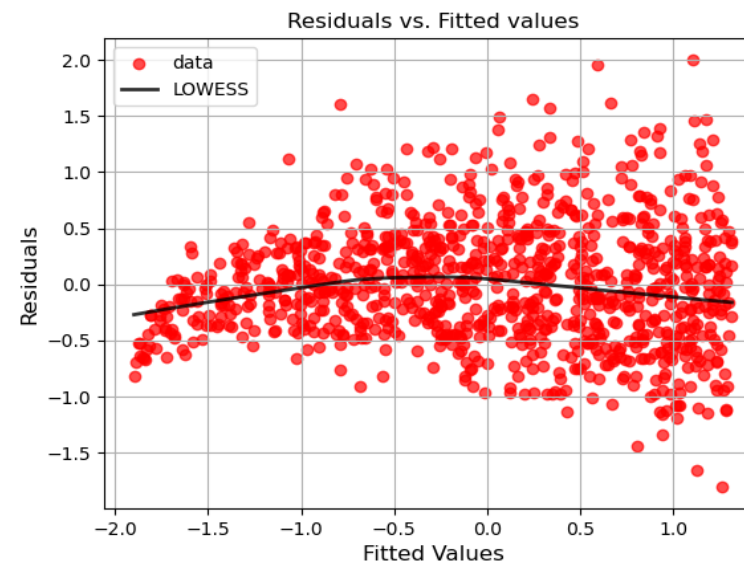
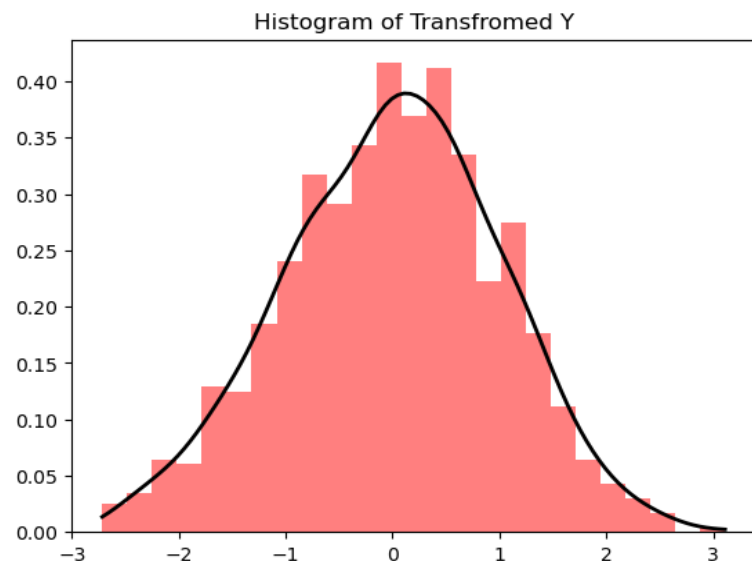
- معمولاً اگر چولگی متغیر پاسخ بالا باشد، احتمال بروز ناهمسانی واریانس افزایش می‌یابد.



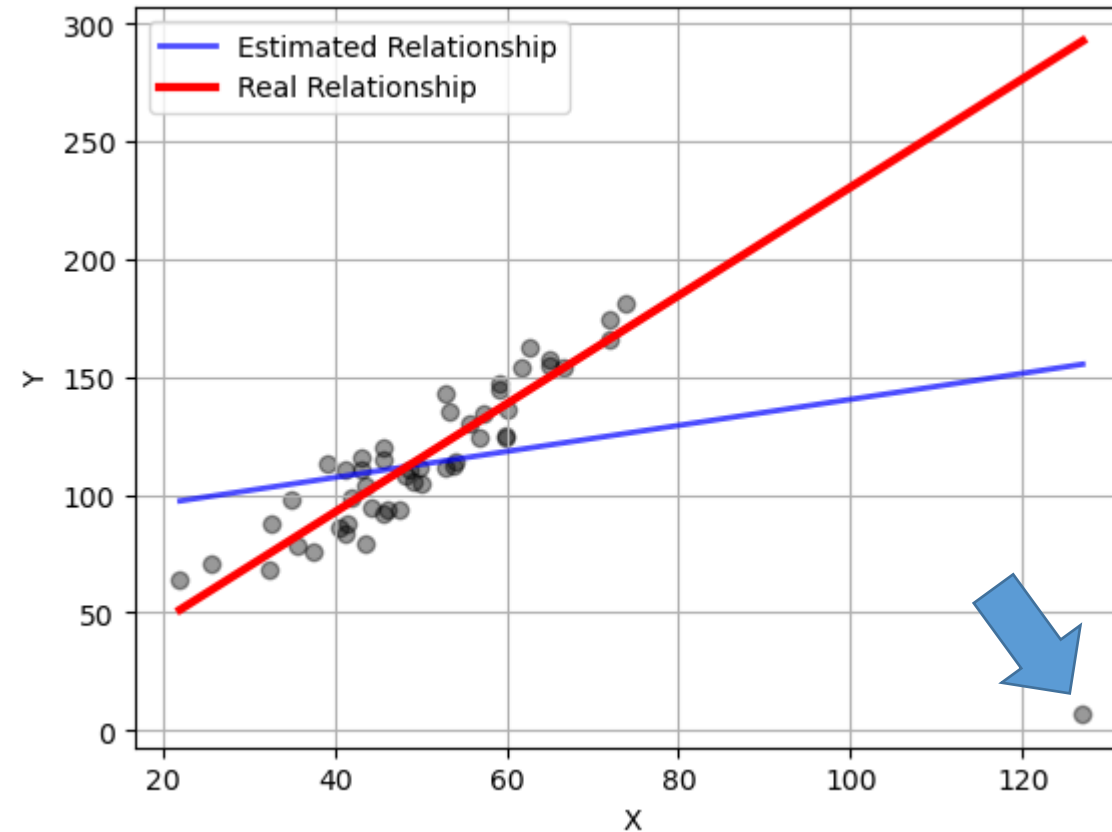
تغییر متغیر Box-Cox

- تغییر متغیر Box-Cox کمک می‌کند تا توزیع متغیر را به توزیع نرمال نزدیک کنیم

$$y_i^{(\lambda)} = \begin{cases} \frac{y_i^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \ln(y_i) & \lambda = 0 \end{cases}$$



داده پرت و مشاهدات اثرگذار بر مدل رگرسیون خطی



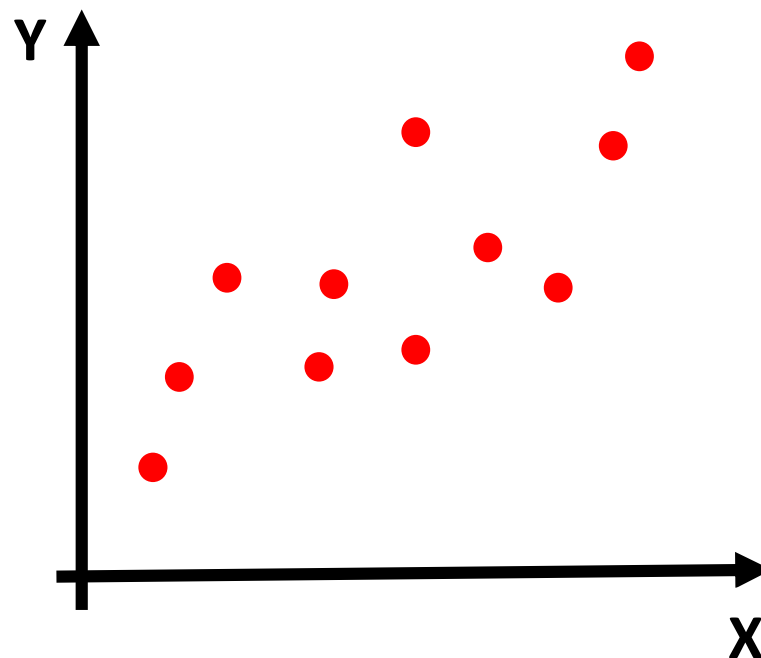
قاعده‌سازی (Regularization) در رگرسیون خطی

• تابع هزینه در Ridge Regression (L2 Regularization)

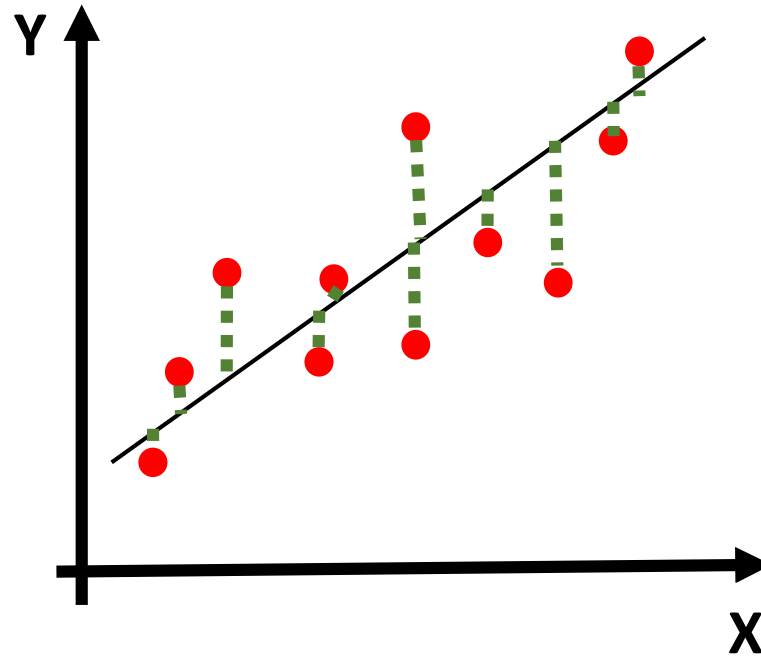
$$\text{Minimize}_{\beta}: \text{MSE} + \frac{\alpha}{n} \sum_{j=1}^m \hat{\beta}_j^2, \alpha \geq 0$$

• هرچه α بزرگتر شود، ضرایب به صفر میل می‌کنند ولی صفر نمی‌شوند.

فهم شهودی رویکرد Ridge Regression

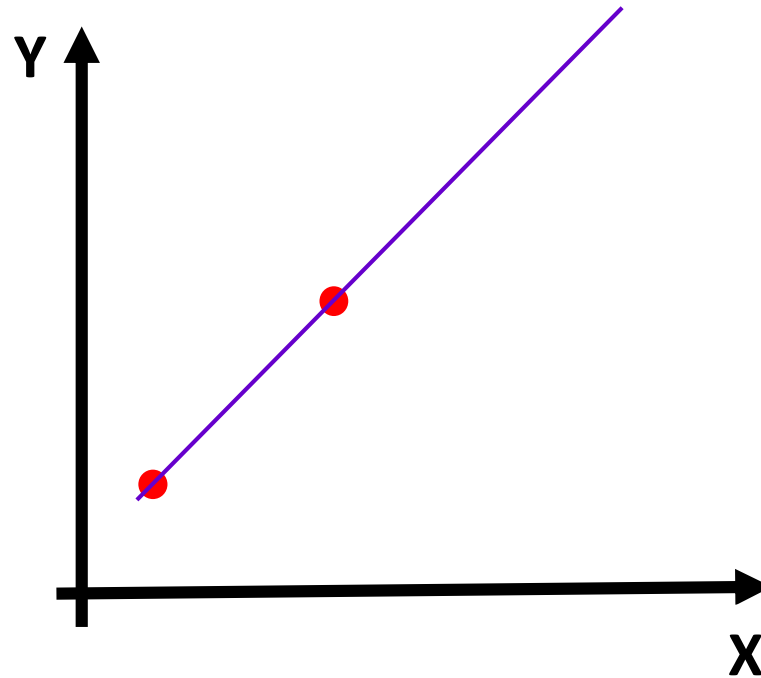


فهم شهودی رویکرد Ridge Regression



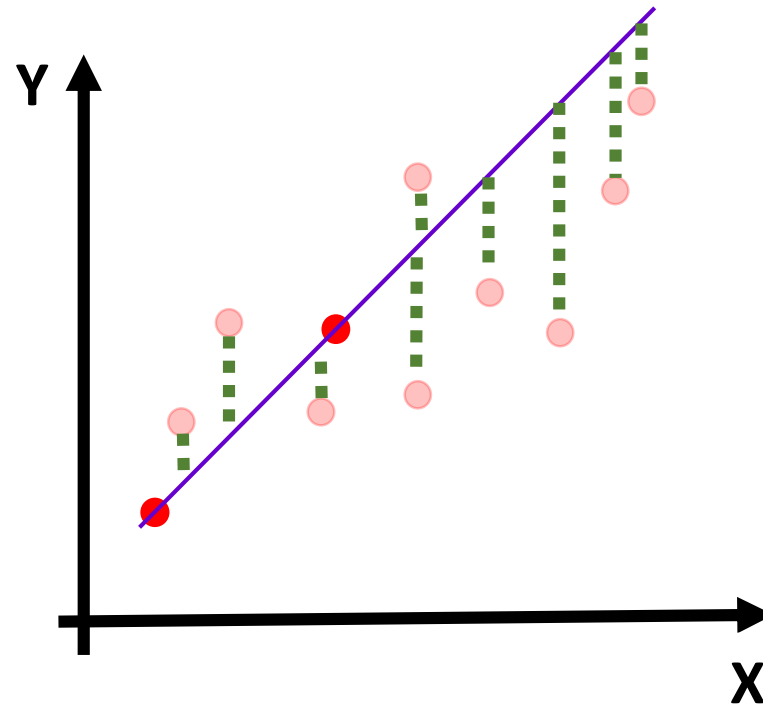
مدل رگرسیون خطی به روش OLS

فهم شهودی رویکرد Ridge Regression



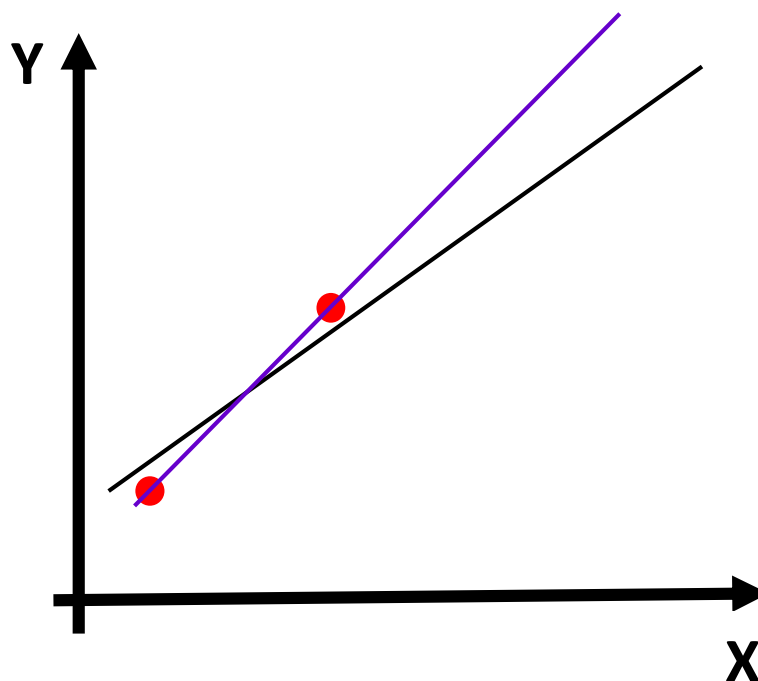
مدل رگرسیون خطی به روش OLS

فهم شهودی رویکرد Ridge Regression



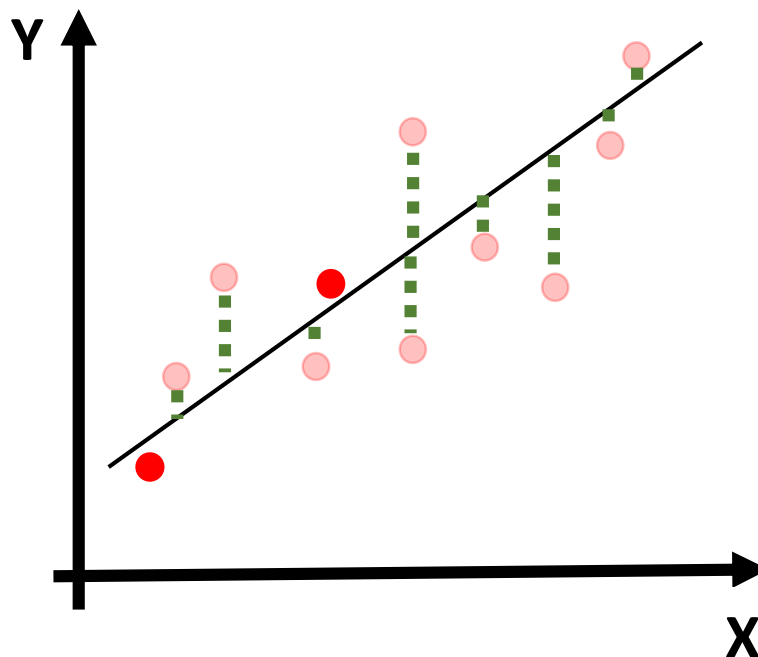
مدل واریانس بالایی دارد

فهم شهودی رویکرد Ridge Regression



مدل Ridge Regression
به مدل کمی سوگیری (Bias) اضافه کنیم

فهم شهودی رویکرد Ridge Regression



با برازش مدلی که روی داده‌های آموزش
سوگیری دارد، واریانس را در داده‌های آزمایش
کاهش می‌دهیم تا به قدرت پیش‌بینی بهتری
برسیم

قاعده سازی (Regularization) در رگرسیون خطی

• تابع هزینه در LASSO Regression (L1 Regularization)

$$\text{Minimize}_{\beta}: MSE + \frac{\alpha}{n} \sum_{j=1}^m |\hat{\beta}_j|, \alpha \geq 0$$

• ضرایب متغیرهای پیش‌بینی‌کننده کم‌اهمیت می‌تواند صفر شود. بنابراین از LASSO می‌توان برای انتخاب ویژگی‌ها استفاده کرد.

فرمول بندی دیگر

• استفاده از رویکرد ضرایب لاگرانژ (Lagrange Multiplier)

$Minimize_{\beta}: MSE$

• Ridge Regression رویکرد

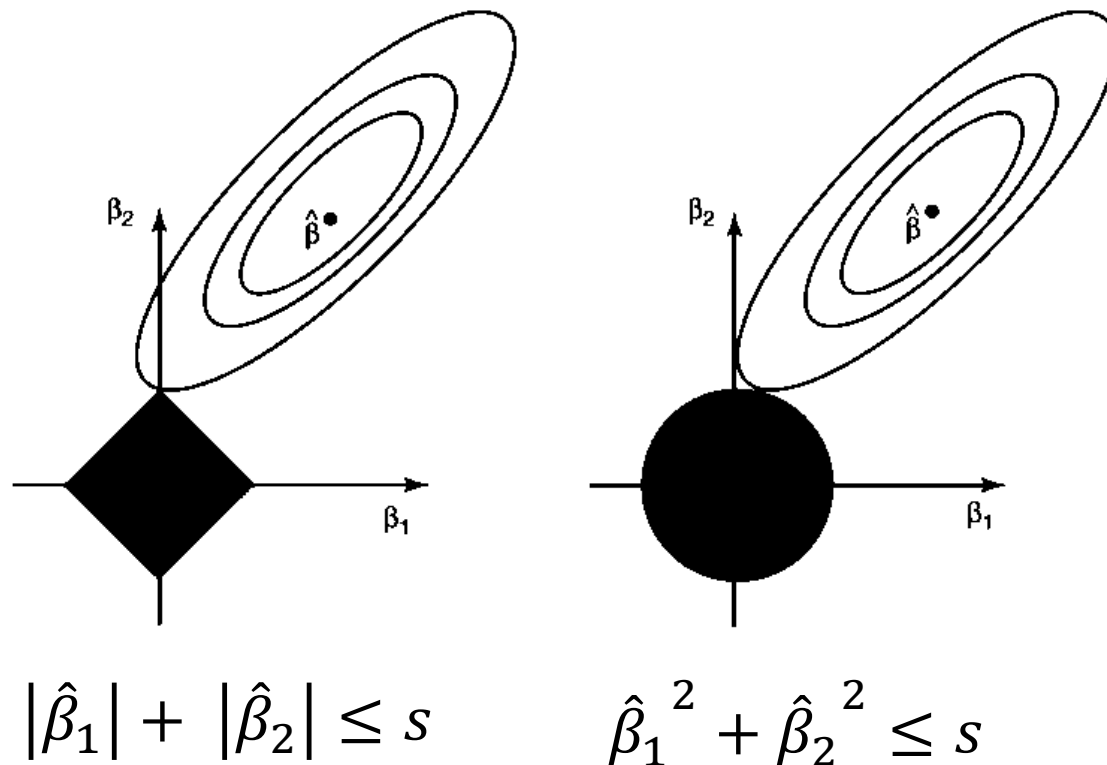
$$Subject\ to: \sum_{j=1}^m \hat{\beta}_j^2 \leq s$$

$Minimize_{\beta}: MSE$

• LASSO Regression رویکرد

$$Subject\ to: \sum_{j=1}^m |\hat{\beta}_j| \leq s$$

مقایسه Ridge Regression و LASSO Regression



چند نکته درباره قاعده‌سازی در رگرسیون خطی

- الگوریتم Ridge Regression توسط کلاس `sklearn.linear_model.Ridge` اجرا می‌شود.
- الگوریتم LASSO توسط کلاس `sklearn.linear_model.Lasso` اجرا می‌شود.
- می‌توان از روش k-Fold CV برای انتخاب فرایارامتر بهینه بهره برد.
- برای استفاده از الگوریتم‌های قاعده‌سازی، نیاز است همه ویژگی‌ها مقیاس شوند.
- وقتی تعداد متغیرهای پیش‌بینی‌کننده زیاد و احتمال بروز هم‌خطی بین آنها وجود دارد، قاعده‌سازی مفید است.
- وقتی تعداد متغیرها بیشتر از تعداد مشاهدات است، می‌توان از قاعده‌سازی استفاده کرد.

برنامه‌نویسی در پایتون

موردکاوی الگوریتم رگرسیون خطی



