Regressão Linear

Introdução

Modelo de regressão Linear

Forma matricial

Interpretação para o caso Bidimensional

Definições sobre o erro

Método dos Mínimos Quadrados

Forma Matricial

Média e variância de LS para B

Porque usar Mínimos Quadrados

Bias e Variância tradeoff

Estratégias de Regularização

Introdução

Na estatística →
- eq. para estimar o valor esperado de y dado x.
- a relação entre a resposta (y) e as entradas é uma relação linear
- São ajustados usando uma abordagem de mínimos quadrados

Extensamente usada em problemas reais, especialmente em IA, em classic modelo

Equação da Regressão Linear

Para estimar o valor de y usa-se o seguinte modelo para o caso de multíplas entradas.

$$y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip} + \varepsilon_i$$

Onde y_i é o valor esperado que o modelo irá prever; β são os coeficientes de ajuste, X_i a variável explicativa (entradas) e ε_i o erro amostral do erro de medição. Ele engloba todos os fatores residuais da estimação.

Nesse caso, podemos dizer que

$$\boxed{\hat{y}_i = y_i - \varepsilon_i}$$

Forma Matricial

$$\boxed{y = X\beta + \varepsilon}$$

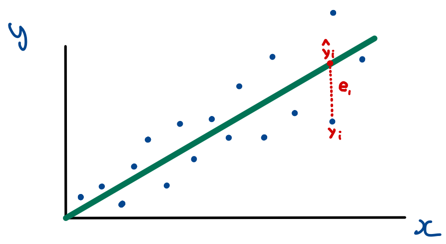
* X é fixo, não é uma V.A.
* β são os parâmetros, não é uma V.A
* y é a variável dependente, é V.A
* e é uma V.A com $E[e_i]=0$ e $Var[e_i] = \rho^2$

$$\underset{m\times 1}{y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}} \quad \underset{m \times p+1}{X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{m1} & \cdots & x_{mp} \end{bmatrix}} \quad \underset{p\times 1}{\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}} \quad \underset{m\times 1}{\varepsilon = \overbrace{\begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_m \end{bmatrix}}^{\text{Resíduos}}}$$

Interpretação para o caso Bidimensional

Por simplicidade, vamos considerar o caso bidimensional, onde:

$$\boxed{y_i = \beta_0 + \beta_1 x_i + e_i}$$



Definições sobre o Erro:

* $E[e_i] = \bar{e}_i = 0$
* $Var[e_i] = E[(e_i - \bar{e}_i)^2] = \sigma^2$
* $E[y_i] = E[(\beta_0 + \beta_1 x_i + e_i)] = \beta_0 + \beta_1 x_i = \mu_i$ ⤳ $\boxed{E[Y] = \beta X}$
* $Var[y_i] = \sigma^2$
* Como assumimos que $e_i \simeq N(0,\sigma^2)$, e_i e e_j $\forall i,j$ $i \neq j$ não independentes, $cov(e_i, e_j) = 0$

Método dos Mínimos Quadrados

* Como estimar $[B_0, B_1, \dots, B_p]^T$?
* Método mais utilizado é o mínimos quadrados (Least squares)
* técnica de otimização matemática que procura encontrar o melhor ajuste (B) para o conjunto de dados que minimize a soma dos quadrados entre o valor estimado ($\hat{y}$) e o valor observado (y)
* Essas diferenças são chamadas de resíduos

$$\boxed{L(B) = \sum_{i=1}^{m} (y_i - E[y_i])^2}$$

Vamos analisar o caso bidimensional:

$$L(B) = \sum_{i=1}^{m} (y_i - E[y_i])^2 = \sum_{i=1}^{m} (y_i - B_0 - B_1 x_i)^2$$

Para minimizar, basta derivar em relação a B_0 e B_1:

$$\frac{\partial L}{\partial B_0} = -2 \sum_{i=1}^{m} (y_i - B_0 - B_1 x_i) = 0$$

$$-2\left[\sum_{i=1}^{m} y_i - mB_0 - B_1 \sum_{i=0}^{m} x_i\right] = 0$$

$$\boxed{B_0 = \bar{y} - B_1 \bar{x}}$$

$$L(B) = \sum_{i=1}^{m} (y_i - B_0 - B_1 x_i)^2$$

$$\begin{cases} \dfrac{\partial L}{\partial B_1} = -2 \sum_{i=1}^{m} (y_i - B_0 - B_1 x_i) x_i = 0 \\ B_0 = \bar{y} - B_1 \bar{x} \end{cases}$$

$$\boxed{B_1 = \frac{\sum_{i=1}^{m} x_i (y_i - \bar{y})}{\sum_{i=1}^{m} x_i (x_i - \bar{x})}}$$

Forma Matricial

$$L(B) = (\underbrace{y - XB}_{})^T (y - XB)$$

$$\square \times \big| = \square$$

* $L(B) = y^T y - 2B^T X^T y + B^T X^T X B$

$$\frac{\partial L(B)}{\partial B} = 0 - 2X^T y + 2X^T X B \rightarrow \frac{\partial L(B)}{\partial B} = 0$$

$$\boxed{B = (X^T X)^{-1} X^T y}$$

* $(X^T X)^{-1}$ é uma matriz singular
  Não tem solução única
* B não é único $\rightarrow \hat{B}$
  $\hat{y} = X\hat{B}$

Variância de B

$$Var[\hat{B}] = Var[(X^T X)^{-1} X^T y]$$
$$= (X^T X)^{-1} X^T \underbrace{Var[y]}_{} \rightarrow Var[\hat{B}] = (X^T X)^{-1} X^T \sigma^2$$

variância do erro

* Podemos usar um estimador de variância para o erro
* $\sigma^2 = E[e^2] \rightarrow \boxed{\hat{\sigma}^2 = \frac{1}{m-(p+1)} \sum_{i=1}^{m} (y_i - \hat{y}_i)^2}$

número de parâmetros B

Porque usar mínimos quadrados?

* $L(B)$ é um estimador BLUE (Best Linear Unbiased Estimator)
* Um bom estimador deve ser unbiased e low variance
* bias and variance trade-off

Pelo teorema de Gauss-Markov se:

* $E[e_i]=0$
* $Var[e_i]=\sigma^2$ para todo i, (constante)
* $Cov(e_i, e_j)=0$ para todo i,j onde $i\neq j$

Então OLS é o melhor estimador BLUE

LS: * $\hat{B} = \overbrace{(X^TX)^{-1}X^T}^{C} Y = C\,E[Y] = C\,E[XB+e]$

* $E[\hat{B}] = \underbrace{C}_{(X^TX)^{-1}X^T}[\underbrace{E[XB]}_{XB} + \underbrace{\cancel{E[e]}}_{0}] = \underbrace{\cancel{(X^TX)^{-1}(X^TX)}}_{I}B$ $\boxed{E[\hat{B}] = IB}$

* $Var[\hat{B}] = Var[(X^TX)^{-1}X^TY] = Var[CY] = C\,\underbrace{\cancel{Var[Y]}}_{\sigma^2}C^T$

$Var[\hat{B}] = CC^T\sigma^2 = (X^TX)^{-1}\sigma^2$

LS tem low Bias mas alto variância

como reduzir? estratégias para adaptar a função custo → Regularização

↑ Bias ↓ variance

Ridge Regression

$$\boxed{L_{ridge}(\hat{B}) = \sum_{i=1}^{N}\left(y_i - B_0 - \sum_{j=1}^{p} x_{ij}B_j\right)^2 + \lambda\sum_{j=1}^{p} B_j^2}$$

* L2 regularization
* λ controla a contração

derivar e igualar a zero

$$\boxed{\hat{B} = (I\lambda + X^TX)^{-1}X^TY}$$

* melhor quando varias entradas
* adiciona mais bias na estimação de B

Regressão Lasso

$$\boxed{L_{ridge}(\hat{B}) = \sum_{i=1}^{N}\left(y_i - B_0 - \sum_{j=1}^{p} x_{ij}B_j\right)^2 + \lambda\sum_{j=1}^{p} |B_j|}$$

* Segura o tamanho de B