

Tarea 3: Regresión lineal (parte teórica)

Sea Y la altura de una población y X una matriz de datos como peso, ancho de huesos, etc. con P variables observadas N veces:

+ Plantea el problema de regresión como un problema de mínimos cuadrados, donde $\hat{\beta}^* = \operatorname{argmin}_{\beta \in \mathbb{R}^p} \|Y - X^T \vec{\beta}\|^2$ donde $\hat{\beta}^* = \hat{\beta}_1^*, \hat{\beta}_2^*, \dots, \hat{\beta}_p^*$

$$\min_{\vec{\beta}} (Y - X^T \vec{\beta})^2 = \min_{\vec{\beta}} (Y - \hat{Y})^2 = \min_{\vec{\beta}} (U)^2$$

+ La solución consiste en resolver las condiciones de primer orden de este problema de minimización i/p¿ i/p¿ Donde de las C.P.O se tiene que: i/p¿

$$XY - XX^T \vec{\beta} = 0$$

$$\Rightarrow \beta^* = [XX^T]^{-1}XY$$

+ La razón por la cual este método da como solución un ajuste lineal a los datos es la siguiente: i/p¿ Recordemos que el modelo bajo el cual estamos asumiendo la relación de \mathbf{Y} con \mathbf{X} es tal que se puede escribir de la siguiente forma (en un modelo centrado) i/p¿ i/p¿

$$\mathbf{Y} = \beta_1 X_i^1 + \beta_2 X_i^2 + \dots \beta_P X_i^P$$

con

$$i \in \{1, \dots, N\}$$

i/p¿ i/p¿ Sí, podríamos usar este método para ajustar polinomios de la forma: $y = x^2$, ya que el modelo es lineal en los parámetros. Por lo que bastaría con definir $\mathbf{x}_i^2 = u_i; \forall i \in \{1, \dots, N\}$ y encontrar el resultado de las ecuaciones que resultan del problema del apartado anterior para encontrar el vector que minimiza la distancia entre \mathbf{Y} y $\hat{\mathbf{Y}}$ + La relación entre la solución encontrada y un problema de proyección en subespacios vectoriales es que $\vec{\beta}^*$ es la proyección de \mathbf{Y} en el espacio generado por las columnas de $\mathbf{X}X^T$, dicha proyección es tal que siempre es ortogonal i/p¿ + La relación con el teorema de pitágoras es que en términos geométricos la recta generada por \vec{Y} junto con la recta que genera la proyección \hat{Y} y el error de proyección \vec{U} con $\mathbf{P} = 2$ describen un triángulo en el cual su Cateto Adyacente (C.A) es igual \hat{Y} , su Cateto Opuesto (C.O) es igual \vec{U} y su Hipotenusa (H) es igual a \hat{Y} . Por lo que si nombramos θ al ángulo formado por C.A y H, su coseno sería igual al cociente definido por la magnitud de H y la magnitud de C.A + Agregar una columna de unos a la matriz de datos favorece la estimación de la proyección al permitir que la proyección \hat{Y} no forzosamente pase por el origen + El problema de regresión desde la estadística generalmente usa una función de pérdida de la forma: $\mathcal{L}(Y|1, X) = (Y - b_0 - X^T \vec{b})^2$ (modelo lineal). Y, busca modelar $\mathbf{E}(Y|X)$, para esto asume una distribución estadística del error del modelo ε . Con base en esta distribución los modelos más conocidos: i/p¿ Primero, el modelo lineal consiste en suponer una relación paramétrica entre la variable dependiente \mathbf{Y} y \mathbf{X} de la forma: i/p¿

$$\mathbf{Y}_i = \beta_0 + \beta_1 X_i^1 + \beta_2 X_i^2 + \dots + \beta_P X_i^P + \epsilon_i$$

¡p! En el modelo lineal: ¡/p!

1) Si $\mathbf{E}(\epsilon|X) = 0$, se le conoce como el Modelo de Regresión Lineal. Es de los menos restrictivos, pues no asume alguna distribución en particular para los errores y tampoco pide que el valor esperado de los errores sea cero (es decir no pide que ϵ sea forzosamente ruido blanco), en su lugar pide que las \mathbf{X} variables explicativas no aporten información a ϵ

2) Si \mathbf{U} es independiente a \mathbf{X} , se le conoce como Modelo Clásico de Regresión Lineal (independencia implica $\mathbf{E}(\epsilon X) = E(\epsilon)E(X)$)

3) Si la distribución condicional de \mathbf{Y} dado \mathbf{X} es Gaussiana con varianza constante, al modelo se le conoce como Modelo de Regresión Lineal Normal (donde $\epsilon \sim N(0, \sigma^2)$)

4) Si $\mathbf{E}(\epsilon X) = 0$, se le conoce como Modelo de Correlación

+ En el Modelo de Regresión Lineal Normal la función de verosimilitud se construye: ¡p! Sea ¡/p!

$$\mathbf{Y}_i = \beta_0 + \beta_1 X_i^1 + \beta_2 X_i^2 + \dots + \beta_P X_i^P + \epsilon_i$$

¡p! en forma vectorial: ¡/p!

$$\Rightarrow \mathbf{Y} = \mathbf{X}^T \beta + \epsilon$$

¡p! con ¡/p!

$$\epsilon \sim N(0, \sigma^2 I_n)$$

$$\Rightarrow \mathbf{Y} \sim N(\mathbf{X}^T \beta, \sigma^2 I_n)$$

¡p! Por lo que la función de verosimilitud en este caso es ¡/p!

$$\Rightarrow L(\beta, \sigma^2; X) = \prod \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(Y - X^T \beta)^2}{2\sigma^2}}$$

+ Para mostrar que la solución al problema de maximizar la función de verosimilitud es igual al problema de mínimos cuadrados del Modelo de Regresión Lineal Normal, es más sencillo si se le aplica el logaritmo a la función de verosimilitud (el logaritmo es una transformación monótona, por lo que preserva el orden) ¡p! ¡/p!

$$\Rightarrow \log(L(\beta, \sigma^2; X)) = -\frac{1}{\sqrt{2\pi\sigma^2}} - \frac{(Y - X^T \beta)^2}{2\sigma^2}$$

¡p! Y la C.P.O ¡/p!
[β]

$$\Rightarrow \mathbf{X}(Y - \mathbf{X}^T \beta) = 0$$

$$\Rightarrow \mathbf{X}Y - \mathbf{X}\mathbf{X}^T \beta = 0$$

$$\Rightarrow \beta = [\mathbf{X}\mathbf{X}^T]^{-1} \mathbf{X}Y$$

+ Teorema de Gauss Markov (de Wikipedia.org) ¡p! Dice que bajo el Modelo de Regresión Lineal, el estimador que resulta del método mínimos cuadrados (OLS) es el mejor estimador insesgado (BLUE) de los coeficientes. Es el mejor en el sentido de que tienen la menor varianza cuando se compara con otros estimadores lineales insesgados ¡/p!