Complemento sobre Esperanza Condicional, Normal bivariada y Regresión lineal

Mathias Bourel

IMERL - Facultad de Ingeniería, Universidad de la República, Uruguay

April 22, 2021

Recordamos que el objetivo del aprendizaje supervisado en un problema de regresión consiste en buscar una función f que minimice el riesgo empírico

$$\mathbb{E}_{(X,Y)}\left[\left(Y - f(X)\right)^2\right]$$

La función f que minimiza esta función es la esperanza condicional $m(X) = \mathbb{E}(Y|X)$. En efecto si $f : \mathbb{R}^d \to \mathbb{R}$ entonces:

$$\mathbb{E}\left[(f(X) - Y))^2\right] = \mathbb{E}\left[(f(X) - m(X) + m(X) - Y))^2\right] = \mathbb{E}\left[(f(X) - m(X))^2\right] + \mathbb{E}\left[(m(X) - Y))^2\right]$$

y es claro que esta expresión es mínima cuando $f(x) = m(x) = \mathbb{E}(Y|X = x)\ \forall\, x$.

En el cálculo anterior usamos que:
$\mathbb{E}_{(X,Y)}\left[(f(X) - m(X) + m(X) - Y))^2\right]$
$= \mathbb{E}\left[(f(X) - m(X))^2\right] + \mathbb{E}\left[(m(X) - Y)^2\right] + 2\mathbb{E}((f(X) - m(X))(m(X) - Y))$
y $\mathbb{E}_{(X,Y)}((f(X) - m(X))(m(X) - Y)) =$
$\mathbb{E}_X(\mathbb{E}((f(X) - m(X))(m(X) - Y)|X)) = \mathbb{E}_X((f(X) - m(X))(\mathbb{E}(m(X) - Y|X)) =$
$\mathbb{E}_X((f(X) - m(X))(\mathbb{E}(Y|X) - \mathbb{E}(Y|X)) = 0.$

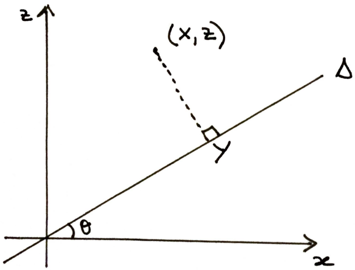
Conclusión:

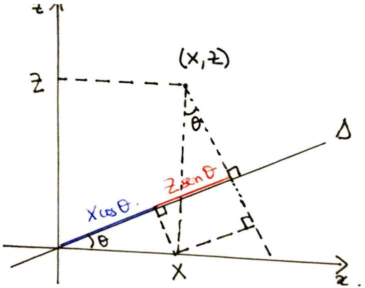
$$\mathbb{E}(Y|X) = \underset{f}{\text{Argmin}}\ \mathbb{E}\left[\left(Y - f(X)\right)^2\right]$$

La función $m(X)$ se conoce también como *regresión* de Y sobre X.

Motivación

Sean X e Z dos variables aleatorias independientes normales estadándares ($X, Z \sim \mathcal{N}(0,1)$). Entonces (X, Z) es un vector aleatorio con distribución normal estándar. Consideramos la recta Δ que pasa por el origen y que forma un ángulo de θ con el eje (Ox). Sea Y la proyección de (X, Z) sobre Δ. En realidad identificamos Y con la norma del vector OY.





Por lo tanto $Y = X\cos\theta + Z\sin\theta$ y entonces al ser combinación lineal de dos normales, Y tiene distribución normal. Más aún $Y \sim \mathcal{N}(0,1)$ ya que:

$$\mathbb{E}(Y) = \mathbb{E}(X\cos\theta + Z\sin\theta) = \cos\theta\mathbb{E}(X) + \sin\theta\mathbb{E}(Z) = 0$$

$$\mathrm{Var}(Y) = \mathrm{Var}(X\cos\theta + Z\sin\theta) = \cos^2\theta\mathrm{Var}(X) + \sin^2\theta\mathrm{Var}(Z) = cos^2\theta + \sin^2\theta = 1$$

El coeficiente de correlación entre X e Y es entonces:

$$\rho = \rho_{X,Y} = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y} = \mathrm{Cov}(X,Y) = \mathbb{E}(XY) = \mathbb{E}(X(X\cos\theta + Z\sin\theta)) = \mathbb{E}(X^2)\cos\theta = \cos\theta$$

Para cada $\rho \in [-1,1]$, existe un ángulo θ tal que $\rho = \cos\theta$. Luego, para cada $\rho \in [-1,1]$ existen X e Y normales estándar con correlacioón ρ. Podemos escribir entonces:

$$Y = \rho X + \sqrt{1-\rho^2} Z$$

en donde X e Z son normal estándares independientes.

Definición:

1. Decimos que el par (X,Y) tiene distribución normal estándar con correlación ρ si existe Z normal estándar independiente de X tal que

$$Y = \rho X + \sqrt{1-\rho^2} Z$$

2. Decimos que el par (X,Y) tiene distribución normal con parámetros $(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho)$ si $\left(\frac{X-\mu_X}{\sigma_X}, \frac{Y-\mu_Y}{\sigma_Y}\right)$ tiene distribución normal estándar con correlación ρ.

De la definición anterior se desprende que existe una variable aleatoria Z normal estándar, independiente de X tal que

$$\frac{Y-\mu_Y}{\sigma_Y} = \rho\frac{X-\mu_X}{\sigma_X} + \sqrt{1-\rho^2}Z$$

es decir:

$$Y = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(X-\mu_X) + \sigma_Y\sqrt{1-\rho^2}Z$$

Veamos como queda en este caso la función de regresión:

$$\begin{aligned}
m(x) =& \mathbb{E}(Y|X=x) = \mathbb{E}\left(\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(X-\mu_X) + \sigma_Y\sqrt{1-\rho^2}Z|X=x\right) \\
=& \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X) + \sigma_Y\sqrt{1-\rho^2}\mathbb{E}\left(Z|X=x\right) \\
=& \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X) + \sigma_Y\sqrt{1-\rho^2}\mathbb{E}\left(Z\right) \\
=& \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X)
\end{aligned}$$

ya que Z es independiente de X. Este razonamiento muestra que $m(x)$ es una función lineal en el caso que (X, Y) sea un vector normal bivariado. Observar que

$$m(x) = \beta_0 + \beta_1 x \quad \text{donde} \quad \beta_0 = \mathbb{E}(Y) - \beta_1\mathbb{E}(X) \quad \text{y} \quad \beta_1 = \frac{Cov(X,Y)}{\sigma_X^2}$$