Tema 4 Parte II. Regresión Mínimo Cuadrática Bidimensional

Doble Grado en Ingeniería Informática y Matemáticas.

Úrsula Torres Parejo

El problema de regresión consiste en aproximar una variable aleatoria (Y) mediante una función de otra variable (X) definida sobre el mismo espacio de probabilidad: $Y \approx \varphi(X)$

- Y: Variable dependiente, explicada o endógena
- X: Variable independiente, explicativa o exógena
- φ: Función de regresión

FUNCIÓN DE REGRESIÓN: El problema consiste en elegir la función de regresión φ de forma que las predicciones realizadas a partir de ella sean óptimas.

Criterio de optimabilidad: Mínimos Cuadrados: Consiste en elegir φ de forma que las desviaciones cuadráticas entre los valores reales y los aproximados sea mínima, esto es:

$$\boxed{\varphi(X) = \text{Min } E[(Y - \varphi(X))^2] = \text{Min E.C.M.}}$$

Error Cuadrático Medio: $\boxed{\text{E.C.M.} = E\,[(Y - \varphi(X))^2]}$

Solución del Problema:

$E[(Y-\varphi(X))^2] = E[E[(Y-\varphi(X))^2]|X] \longrightarrow$

$\text{Min } E[(Y-\varphi(X))^2] = \text{Min } E[E[(Y-\varphi(X))^2]|X]$

Por propiedad vista para variables aleatorias $E[X-c]^2$ es mínima para $c=E[X]$

$\Rightarrow \boxed{\varphi(X) = E[Y|X]}$

$E.C.M. = E[(Y-\varphi(X))^2] =$

$E[(Y-E[Y|X])^2] =$

$E[E[(Y-E[Y|X])^2]|X] =$

$E[Var[Y|X]] \Rightarrow \boxed{E.C.M. = E[Var[Y|X]]}$

Desarrollando esta última expresión:

$E[Var[Y|X]]= E[E[Y^2|X]-E[Y|X]^2]$

$= E[Y^2] – E[E[Y|X]^2]$

Teniendo en cuenta la descomposición de la varianza:

$Var [Y] = E[Var[Y|X]] + Var[E[Y|X]]$

$= E.C.M. [\varphi(X)] + Var[\varphi(X)]$

$Var[Y] > E[Var[Y|X]]$

Curvas de regresión mínimo cuadráticas

Curva de regresión de Y sobre X

$$Y=E[Y|X=x_0]$$

E.C.M. asociado a la curva=$E[Var[Y|X]]$

E.C.M. para un valor concreto x_0=$Var[Y|X=x_0]$

Curva de regresión de X sobre Y

$$X=E[X|Y=y_0]$$

E.C.M. asociado a la curva=$E[Var[X|Y]]$

E.C.M. para un valor concreto y_0=$Var[X|Y=y_0]$

Se tiene, por tanto:

	Sin observar X	**A partir de X**	**A partir de un valor concreto x_0**
Predicción Mínimo Cuadrática de Y	$E[Y]$	$E[Y\|X]$	$E[Y\|X=x_0]$
Error Cuadrático Medio	$Var[Y]$	$E[Var[Y\|X]]$	$Var[Y\|X=x_0]$

Casos particulares:

1. X e Y independientes → Las curvas de regresión son paralelas a los ejes, la predicción es constante.
   $X = E[X]$, E.C.M. $= Var[X]$
   $Y=E[Y]$, E.C.M. $= Var[Y]$

2. Y depende funcionalmente con X (o viceversa) → La curva de regresión de Y sobre X coincide con la curva de dependencia. $Y = f(X)$, E.C.M. $=0$ (o viceversa)

3. Hay dependencia funcional recíproca → Ambas curvas coinciden y coinciden con la dependencia.
   $Y=f(X)$, E.C.M. $=0$, $X=f^{-1}(Y)$, E.C.M.$=0$

Razón de correlación

Nos da la proporción de la varianza de la variable dependiente explicada por la regresión, o lo que es lo mismo, la concentración en torno a la curva de regresión.
También se conoce como **bondad del ajuste.**

$$\boldsymbol{\eta^2_{Y|X} = \frac{Var[E[Y|X]]}{Var[Y]} = 1 - \frac{E[Var[Y|X]]}{Var[Y]}}$$

$$\boldsymbol{\eta^2_{X|Y} = \frac{Var[E[X|Y]]}{Var[X]} = 1 - \frac{E[Var[X|Y]]}{Var[X]}}$$

Propiedades:

1. $0 \leq \eta^2_{Y|X} \leq 1$
2. $\eta^2_{Y|X} = 0 \Longrightarrow$ Y = E[Y]
3. $\eta^2_{Y|X} = 1 \Longrightarrow$ Y depende funcionalmente de X

   $\eta^2_{X|Y} = 1 \Longrightarrow$ X depende funcionalmente de Y

   $\eta^2_{Y|X} = \eta^2_{X|Y} = 1 \Longrightarrow$ Dependencia funcional recíproca

Regresión Lineal Mínimo Cuadrática

Recta de regresión de Y sobre X

La función de regresión buscada será la recta:

$$\boxed{Y = aX + b.}$$

Los coeficientes a y b que minimizan el error cuadrático medio son:

$$a = \frac{Cov[X,Y]}{Var[X]} \quad b = E[Y] - aE[X]$$

a = $\gamma_{Y|X}$: $Coeficiente\ de\ regresión\ de\ Y\ sobre\ X$

Recta de regresión de X sobre Y

La función de regresión buscada será la recta:

$$\boxed{X = a'Y + b'.}$$

Los coeficientes a' y b' que minimizan el error cuadrático medio son:

$$a' = \frac{Cov[X,Y]}{Var[Y]} \quad b' = E[X] - a'E[Y]$$

a' = $\gamma_{X|Y}$: $Coeficiente\ de\ regresión\ de\ X\ sobre\ Y$

Propiedades de las curvas y rectas de regresión:

1. Si la curva de regresión es una recta $\Rightarrow$ Coincide con la recta de regresión
2. Las rectas de regresión se cortan en el punto (E[X],E[Y])
3. Los coeficientes de regresión coinciden en signo y coinciden en signo con la covarianza
4. Si Cov[X,Y] = 0 $\Rightarrow$ Las rectas vienen dadas por Y = E[Y], X = E[X]
5. Si X e Y están linealmente relacionadas (Si lo está Y con X, también lo estará X con Y) $\Rightarrow$ Las rectas de regresión coinciden con la recta de dependencia.

El Error Cuadrático Medio Asociado a la Predicción de Y por X viene dado por:

$$\boxed{E.C.M. = Var[Y] - \frac{Cov[X,Y]^2}{Var[X]}}$$

Coeficiente de determinación lineal

Proporción de varianza de cada una de las variables debida a la regresión lineal.

$$\boxed{\rho_{X,Y}^2 = \frac{Cov[X,Y]^2}{Var[X]Var[Y]}}$$

Propiedades:

1. $0 \leq \rho_{X,Y}^2 \leq 1$
2. $\rho_{X,Y}^2 = \gamma_{Y|X} \cdot \gamma_{X|Y}$
3. $\rho_{X,Y}^2 = 0$ $\Rightarrow$ Las rectas de regresión son paralelas a los ejes: Y = E[Y], X = E[X]
4. $\rho_{X,Y}^2 = 1$ $\Rightarrow$ Existe dependencia funcional lineal. Las dos rectas coinciden.
5. $0 \leq \rho_{Y,X}^2 \leq \eta_{Y|X}^2 \leq 1$

   $0 \leq \rho_{X,Y}^2 \leq \eta_{X|Y}^2 \leq 1$

   Se da la igualdad si curvas y rectas coinciden.

Coeficiente de correlación lineal

Indica el sentido y el grado de correlación entre X e Y.

$$\boxed{\rho_{X,Y} = \sqrt{\frac{Cov[X,Y]^2}{Var[X]Var[Y]}}}$$

Propiedades:

1. -1 ≤ $\rho_{X,Y}$ ≤ 1
2. Coincide en signo con la covarianza y con los coeficientes de regresión
3. $\rho_{X,Y}$ = 0 $\Rightarrow$ Variables incorreladas: Y = E[Y], X = E[X]
4. $|\rho_{X,Y}|$ = 1 $\Rightarrow$ Dependencia lineal funcional.