Primera Parte

1. Plantear el problema de regresión como un problema de mínimos cuadrados, encontrar el vector ¶

$\hat{\beta} = [\hat{\beta}_1, \ldots, \hat{\beta}_p]^T$ que resuelva $\hat{\beta} = argmin_{\beta \in \mathbb{R}} = ||Y - X\beta||^2$

Dado que el error de estimación $\hat{u}_i$ lo definimos como $y_i - \beta\hat{X}_i$, el problema de mínimos cuadrados que se plantea para minimizar el error de estimación es $\underset{\hat{\beta}}{Min}\ e = \sum_{i=1}^{p} \hat{u}_i^2 = \underset{\hat{\beta}}{Min} \sum_{i=1}^{p} (y_i - \beta X_i)$

Y cuando se deriva e iguala a cero se obtiene como solución

$\beta = \frac{\sum_{i=1}^{p}(y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{p}(x_i - \bar{x})^2}$

¿Por qué este planteamiento nos da un ajuste lineal a nuestros datos? ¿Podríamos usarlo para ajustar polinomios?

Nos da un ajuste lineal porque la derivada de la suma de cuadrados es lineal. Se puede utilizar mínimos cuadrados para ajustar polinomios, pero la función de error que minimizaríamos incluiría alguna β con algún término polinomial

Argumentar la relación entre la solución encontrada y un problema de proyección en subespacios vectoriales de álgebra lineal. ¿Cuál es la relación particular con el teorema de Pitágoras?

Como el problema al que le queremos encontrar una solución es del tipo $Ax = b$, se puede pensar que Ax es una aproximación a b. Entonces el problema general de mínimos cuadrados es encontrar una x tal que haga la distancia $||b - Ax||$ tan pequeña como sea posible. El problema de mínimos cuadrados se puede plantear como un problema de proyección, es decir, como encontrar un vector x que haga que Ax sea el punto en el espacio columna $ColA$ más cercano a b.

$$\hat{b} = proy_{ColA} b$$

Si b está en el espacio columna de A, la ecuación $Ax = \hat{b}$ tiene solución y entonces existe un vector $\hat{b}$ en $\mathbb{R}^n$ tal que

$$A\hat{x} = \hat{b}$$

Esta solución proviene de resolver esta proyección de una relación pitagórica en donde buscamos la distancia mínima en el plano Ax que es la proyección de b y que minimiza el error. Si p es la proyección y e el error, entonces la solución descrita proviene de la relación pitagórica

$$||Ax - b||^2 = ||Ax - p||^2 + ||e||^2$$

¿Qué logramos al agregar una columna de unos en la matriz ?

Al agregar el vector columna de unos a la matriz, lo que estamos haciendo es incluir la intercepta en el problema de mininimización de cuadrados, y entonces estaríamos estimando tanto el valor de la intercepta, como el de la pendiente de la recta que mejor ajusta los datos (en caso de una regresión univariada).

¿Cuál es la función de verosimilitud del problema anterior?

Se escribe la verosimilitud como $L(\beta, \sigma^2) = f(Y|\beta, \sigma^2, X)$, y la función que se maximiza como

$$\prod_{i=1}^{p} f(Y \mid \beta, \sigma^2, X)$$

$$= \frac{1}{(2\Pi\sigma^2)} exp\left\{ \frac{-1}{(2\sigma^2)} \sum_{i=1}^{p} (Y_i - \beta_0 - \beta_1 X_i)^2 \right\}$$

Y al tomar logaritmos de la función de máxima versimilitud tenemos

$$-\frac{n}{2} log(2\pi) - \frac{n}{2} log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{p} (Y_i - \beta_0 - \beta_1 X_i)^2$$

Cuando derivamos respecto a β_0, β_1 y σ^2 e igualamos a cero, encontramos los valores para estos parámetros resolviendo tres ecuaciones

$$(1) \quad \sum_{i=1}^{p} (Y_i - \beta_0 - \beta_1 X_i) = 0$$

$$(2) \quad \sum_{i=1}^{p} x_i (Y_i - \beta_0 - \beta_1 X_i) = 0$$

$$(3) \quad \sum_{i=1}^{p} (Y_i - \beta_0 - \beta_1 X_i)^2 = n\sigma^2$$

Los parámetros β_0, β_1 y σ^2 que son solución a este sistema de tres ecuaciones son:

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{p} (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{p} (x_i - \bar{x})^2}$$

y

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^{p} (Y_i - \beta_0 - \beta_1 X_i) = \sum_{i=1}^{p} e_i$$

Estos son los mismos estimadores que obtuvimos en el problema de mínimos cuadrados

El teorema de Gauss-Markov establece que si se cumplen ciertas condiciones, los estimadores de Mínimos Cuadrados Ordinarios son los mejores estimadores lineales insesgados (BLUE). Las condiciones que se deben cumplir son:

1) Linerariedad en los parámetros

2) La muestra es aleatoria

3) Esperanza condicional del error es cero (El modelo está correctamente especificado y no hay problemas como variables omitidas o causalidad inversa)

4) Varianza es constante

5) No colinearidad perfecta

Si se cumplen estas condiciones, entonces los estimadores de MCO son: a) Insesgados ($E(\hat{\beta}_0 = \beta_0 \;\; y \; E(\hat{\beta}_1 = \beta_1))$; y b) Los mejores ($\hat{\beta}_0 \quad \hat{\beta}_1$ tienen la menor varianza entre todos los demás estimadores lineales insesgados)

Segunda parte

¿Qué tan bueno fue el ajuste?

El ajuste es relativamente bueno, pues el modelo explica 87% de la variabilidad de los datos. Es decir, estas variables independientes explican 87% de la variación en precio. Por otro lado, la grafica de residuales muestra que la mayoria esta centrada en el cero, lo que indica que el modelo esta ajustando bien los datos.

¿Qué medida puede ayudarnos a saber la calidad del ajuste? ¿Cuál fue el valor de que ajustó su modelo y que relación tiene con la calidad del ajuste

Una medida que puede ayudarnos es el error estandar de los residuos. Esta medida nos dice que tanto se desvia nuestra variable de interes (precio) de el ajuste que estamos estimando. En este caso, dado que el valor promedio de todos los diamantes de la muestra es 9,285 (la intercepta), y tenemos un error estandar de 1,393, entonces el porcentaje de error representa aproximadamente 15%; es decir, en promedio cualquier predicion estara fuera del valor real en un 15%.

¿Cuál es el ángulo entre y y su estimacion? Hint: usen la R cuadrada y el arcocoseno.