Tarea 3. Regresión Lineal Parte Teórica

Explicar Y = Altura

p variables $X^1,...X^p$ (peso, ancho de huesos, etc.)

Muestra de *N* individuos, cada variable está representada por un vector de tamaño N

La información de las variables explicativas se puede juntar en una matriz:
$X = [X^1|..|X^p]$ de tamaño nxp

a) Plantear el problema de regresión como un problema de mínimos cuadrados, encontrar el vector que $\hat{\beta} = [\hat{\beta}_1,\dots,\hat{\beta}_p]^T$ que resuelva $\hat{\beta}$ = argmin $||Y - X\beta||^2$ y encontrar la solución teórica.

$F(X) = Y = X\beta + \epsilon$

min f(X)
s.a. g(x) =0

La solución debe cumplir
$\nabla f(x) = \lambda \nabla g(x)$
g(x) =0

$\mathcal{L}(x, \lambda) = f(x) - \lambda \nabla g(x)$
$\nabla \mathcal{L}(x, \lambda) = 0$

La solución:

$(X^TX)\,\hat{\beta} = X^TY$
$\hat{\beta} = (X^TX)^{-1}X^TY$

¿Por qué el planteamiento nos da un ajuste lineal a nuestros datos?
Esto se debe a que el modelo involucra una combinación lineal de los parámetros.

¿Podríamos usarlo para ajustar polinomios (ej, $y=x^2$)
Si es posible, ya que en la vecindad de un punto el comportamiento de un polinomio aproximado por el método de mínimos cuadrados nos dará un ajuste adecuado.

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?

AX se encuentra en el espacio columna y dentro de este plano buscamos por el punto más cercano a b. El punto más cercano es la proyección p en el espacio columna. Adicionalmente, AX – b en el espacio columna es perpendicular al error. Por esta razón podemos utilizar el teorema de Pitágoras para un triangulo recto (c2 = a2 +b2), esto es $||Ax - b||^2 = ||Ax - p||^2 + ||e||^2$

¿Qué logramos al agregar una columna de unos en la matriz X?
Es posible lograr un mejor ajuste del modelo.

Plantear el problema de regresión ahora como un problema de estadística

$Y_i = \beta_0 + \beta_1 X_i^1 + \dots + \beta_p X_i^1 + \epsilon$
$\epsilon_i \sim (0, \sigma^2)$

$E(Y_i) = \beta_0 + \beta_1 X_i^1 + \dots + \beta_p X_i^1$

Min $\sum_{I=1}^{n} (y_i - \hat{y}_i)^2$

¿Cuál es la función de verosimilitud?
$Y = X\beta + \epsilon$
$\epsilon_i \sim (0, \sigma^2 I_n)$

$\mathcal{L}(\lambda; X_1, \dots, X_n) = f(X_1, \dots, X_n \mid \lambda) = \prod_{i=1}^{n} f(X_i \mid \lambda)$
Aplicando el ln
$\hat{\mathcal{L}} = \text{Ln} \frac{1}{n} \text{Ln } \mathcal{L}$
$\hat{\lambda} = \arg\max \hat{\mathcal{L}} (\lambda; X_1, \dots, X_n)$

$\nabla \hat{\mathcal{L}} (\hat{\lambda} \mid x) = 0$

Teorema de Gauss-Markov sobre mínimos cuadrados

El Teorema de Gauss-Markov muestra que el estimador $\hat{\beta}$ de mínimos cuadrados es una buena opción siempre y cuando los errores no estén correlacionados. En el caso de que los errores estén correlacionados existirán mejores estimadores. En general, se puede decir que existen tres casos en los que MLS no será necesariamente el mejor método: 1) cuando los errores estén correlacionados; 2) Cuando la distribución del error tiene una cola larga; y 3) Cuando los predictores están altamente correlacionados.