Regresión Bayesiana:

1.1 Objetivo:

En este proyecto se busca que el estudiante emplee la inferencia Bayesian para describir un modelo de regresión lineal y se compare con el modelo puntual basado en máxima verosimilitud [Bishop, 2007].

1.2 Descripción Modelo:

Consideramos una variable aleatoria y como una función de una variable vectorial $x \in \mathbb{R}^k$. Esto es modelado con una relación lineal, con coeficiente β_j, sumado un ruido Gaussiano i.i.d.

$$y_n = x_{n1}\beta_1 + x_{n2}\beta_2 + \ldots + x_{nK}\beta_K + \epsilon_n, \quad \epsilon \sim \mathcal{N}(0, \sigma^2).$$

De forma matricial se representa

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \ldots & x_{1k} \\ x_{21} & x_{22} & \ldots & x_{2k} \\ \vdots & & \ddots & \vdots \\ x_{N1} & x_{N2} & \ldots & x_{Nk} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_N \end{bmatrix}.$$

O simplemente

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

con $\mathbf{y} \in \mathbb{R}^N$, $\boldsymbol{\beta} \in \mathbb{R}^K$ y $\epsilon \in \mathbb{R}^N$. Es usual asumir que la primera columna de $\mathbf{X}$ es constante formadas por unos, así el termino β_1 representa el intercepto.

1.3 Mínimos cuadrados:

En clase se demostró que el vector $\boldsymbol{\beta}$ se puede estimar puntualmente por máxima verosimilitud como

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \mathbf{X}^\top \mathbf{y},$$

donde la inversa corresponde a la pseudo inversa de Moore-Penrose.

1.4 Regresión Bayesiana Lineal:

Uno de los mayores problemas de la regresión lineal es que lo coeficientes β_k pueden llegar a tomar valores muy grandes (sobreentramiento). Para solventar dicho problema se puede emplear una distribución prior para que los coeficientes no sean tan grandes. Entonces, se introduce el prior conjugado Gaussiano, $\boldsymbol{\beta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Lambda}^{-1})$. Aqui estamos parametrizando la Gaussiana usando la inversa de la matriz de covarianza, o matriz de precisión $\boldsymbol{\Lambda}$, con el fin de realizar los cálculos de manera mas fácil. Una forma comun es seleccionar $\boldsymbol{\Lambda} = \lambda \mathbf{I}$, ($\mathbf{I}$ es la matriz identidad) para un escalar positivo λ.
Ahora el posterior para $\boldsymbol{\beta}$ es

$$p(\boldsymbol{\beta}|\mathbf{y}; \mathbf{X}, \sigma^2) \propto \exp\left(-\frac{1}{2\sigma^2}\left(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right)^\top \left(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right) - \frac{1}{2}\boldsymbol{\beta}^\top \boldsymbol{\Lambda}\boldsymbol{\beta}\right).$$

A partir de la expresión anterior se encuentra que la distribución posterior de $\boldsymbol{\beta}$ esta dado por

$$\boldsymbol{\beta} \sim \mathcal{N}\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right),$$

donde

$$\boldsymbol{\mu} = \left(\mathbf{X}^{\top}\mathbf{X} + \sigma^2\boldsymbol{\Lambda}\right)^{-1}\mathbf{X}^{\top}\mathbf{y},$$
$$\boldsymbol{\Sigma} = \sigma^2\left(\mathbf{X}^{\top}\mathbf{X} + \sigma^2\boldsymbol{\Lambda}\right)^{-1}.$$

Observar que se debe asumir conocido el parámetro σ.

Bases de datos

Para este proyecto se empleará la base de datos Swiss, disponible en `https://raw.githubusercontent.com/vincentarelbundock/Rdatasets/master/csv/datasets/swiss.csv`.

Procedimiento

Para realizar la regresión lineal sobre el conjunto de datos se recomienda seguir los siguientes pasos.

- Realizar un análisis de cada variables de la base de datos por medio de Análisis exploratorio de datos (EDA).
- Se quiere calcular la relación entre *Fertility* ($\mathbf{y}$) con respecto a las variables *agri*, *exam*, *edu*, *catholic* y *infant_mort* ($\mathbf{X}$).
- Realizar la estimación de los parámetros $\boldsymbol{\beta}$ usando Mínimos cuadrados.
- Realizar la estimación del posterior de los parámetros $\boldsymbol{\beta}$ usando regresión Bayesiana lineal (asumir diferentes valores para σ y λ)
- Comparar ambos métodos de estimación, a) valor puntual vs media de la distribución del posterior, b) Graficando el valor puntual de cada parámetro (linea vertical) vs distribución marginal del posterior de cada parámetro.

Referencias

[Bishop, 2007] Bishop, C. M. (2007). *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer, 1 edition.