EST-46114 Métodos Multivariados y Datos Categóricos **ITAM**

Lista de Ejercicios (Parte 1)

Prof.: Juan Carlos Martínez-Ovando *12 de febrero de 2016*

1. Considera el modelo normal multivariado normal para $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ vectores p-dimensionales, con

$$\boldsymbol{x}_i | \boldsymbol{\mu}, \lambda \sim \mathrm{N}_p(\boldsymbol{x} | \boldsymbol{\mu}, \lambda \boldsymbol{I}), \tag{1.1}$$

donde $\boldsymbol{\mu}$ es un vector p-dimensional, λ es un escalar positivo y $\boldsymbol{I}$ es la matriz identidad de dimensión $p \times p$. Interpreta $\lambda \boldsymbol{I}$ como la matriz de precisión del modelo.

Considera adicionalmente que la distribución inicial para $\boldsymbol{\mu}$ y λ está dada por

$$\pi(\boldsymbol{\mu}, \lambda) = \mathrm{N}_p(\boldsymbol{\mu} | m_0, \lambda S_0) \text{Ga-Inv}(\lambda | a_0, b_0), \tag{1.2}$$

con m_0 un vector p-dimensional, S_0 una matriz simétrica positivo definida de dimensión $p \times p$, a_0 y b_0 escalares positivos, todos conocidos.

   a) Muestra que la distribución (1.2) es conjugada para (1.1).

   b) Calcula la distibución inicial predictiva el modelo, dada por

$$p(\boldsymbol{x}) = \int p(\boldsymbol{x} | \boldsymbol{\mu}, \lambda) \pi(\boldsymbol{\mu}, \lambda) \mathrm{d}\boldsymbol{\mu} \mathrm{d}\lambda. \tag{1.3}$$

   c) Calcula la distibución posterior para $\boldsymbol{\mu}$ y λ dado $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$.

   d) Calcula la distibución predictiva para $\boldsymbol{x}_f$, un valor futuro no observado aun, dado $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$.

2. Considera el modelo lineal de regresión para $(y_1, \boldsymbol{x}_1), \ldots, (y_n, \boldsymbol{x}_n)$, para $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ vectores p-dimensionales, expresado en forma matricial como

$$\boldsymbol{y} | \boldsymbol{X}, \boldsymbol{\beta}, \lambda \sim \mathrm{N}_n(\boldsymbol{y} | \boldsymbol{X}\boldsymbol{\beta}, \lambda \boldsymbol{I}), \tag{1.4}$$

donde $\boldsymbol{y} = (y_1, \ldots, y_n)'$ es un vector n-dimesional, $\boldsymbol{X} = (\boldsymbol{x}'_1, \ldots, \boldsymbol{x}'_n)'$ es una matriz de dimensión $n \times p$, $\boldsymbol{\beta}$ es un vector p-dimensional, λ es un escalar y $\boldsymbol{I}$ es la matriz identidad de dimensión $n \times n$. De nuevo, interpreta $\lambda \boldsymbol{I}$ como la matriz de precisión del modelo.

Considera adicionalmente que la distribución inicial para $\boldsymbol{\beta}$ y λ está dada por

$$\pi(\boldsymbol{\beta}, \lambda) = \mathrm{N}_p(\boldsymbol{\beta} | b_0, \lambda S_0) \text{Ga-Inv}(\lambda | c_0, d_0), \tag{1.5}$$

con b_0 un vector p-dimensional, S_0 una matriz simétrica positivo definida de dimensión $p \times p$, c_0 y d_0 escalares positivos, todos conocidos.

a) Muestra que la distribución (1.5) es conjugada para (1.4).

b) Calcula la distibución inicial predictiva el modelo, dada por

$$p(\boldsymbol{y} | \boldsymbol{X}) = \int p(\boldsymbol{y} | \boldsymbol{X}, \boldsymbol{\beta}, \lambda) \pi(\boldsymbol{\beta}, \lambda) \mathrm{d}\boldsymbol{\beta} \mathrm{d}\lambda. \tag{1.6}$$

c) Calcula la distibución posterior para $\boldsymbol{\beta}$ y λ dado $(y_1, \boldsymbol{x}_1), \ldots, (y_n, \boldsymbol{x}_n)$.

d) Calcula la distibución predictiva para y_f, un valor futuro no observado aun de y, dado $(y_1, \boldsymbol{x}_1), \ldots, (y_n, \boldsymbol{x}_n)$ y $\boldsymbol{x}_f$ observado.

3. Considera el modelo lineal de regresión para $(y_1, \boldsymbol{x}_1), \ldots, (y_n, \boldsymbol{x}_n)$, para $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ vectores p-dimensionales, expresado en forma matricial como

$$\boldsymbol{y} | \boldsymbol{X}, \boldsymbol{\beta}, \lambda \sim \mathrm{N}_n(\boldsymbol{y} | \boldsymbol{X}\boldsymbol{\beta}, \lambda \boldsymbol{I}), \tag{1.7}$$

donde $\boldsymbol{y} = (y_1, \ldots, y_n)'$ es un vector n-dimesional, $\boldsymbol{X} = (\boldsymbol{x}'_1, \ldots, \boldsymbol{x}'_n)'$ es una matriz de dimensión $n \times p$, $\boldsymbol{\beta}$ es un vector p-dimensional, λ es un escalar y $\boldsymbol{I}$ es la matriz identidad de dimensión $n \times n$. De nuevo, interpreta $\lambda \boldsymbol{I}$ como la matriz de precisión del modelo.

Considera adicionalmente que la distribución inicial para $\boldsymbol{\beta}$ y λ está dada por

$$\pi(\boldsymbol{\beta}, \lambda) = \mathrm{N}_p(\boldsymbol{\beta} | b_0, S_0) \text{Ga-Inv}(\lambda | c_0, d_0), \tag{1.8}$$

con b_0 un vector p-dimensional, S_0 una matriz simétrica positivo definida de dimensión $p \times p$, c_0 y d_0 escalares positivos, todos conocidos.

a) Responde argumentando si la distribución (1.8) es conjugada para (1.7).

b) Calcula la distibución inicial predictiva el modelo, dada por

$$p(\boldsymbol{y}|\boldsymbol{X}) = \int p(\boldsymbol{y}|\boldsymbol{X}, \boldsymbol{\beta}, \lambda)\pi(\boldsymbol{\beta}, \lambda)\mathrm{d}\boldsymbol{\beta}\mathrm{d}\lambda. \tag{1.9}$$

c) Calcula la distibución posterior para $\boldsymbol{\beta}$ y λ dado $(y_1, \boldsymbol{x}_1), \ldots, (y_n, \boldsymbol{x}_n)$.

d) Calcula la distibución predictiva para y_f, un valor futuro no observado aun de y, dado $(y_1, \boldsymbol{x}_1), \ldots, (y_n, \boldsymbol{x}_n)$ y $\boldsymbol{x}_f$ observado.

4. Para el modelo de regresión de la pregunta 1 se sabe que tanto la distribución final para $(\boldsymbol{\beta}, \lambda)$ como la distribución predictiva de y_f en $\boldsymbol{x}_f$, dado los datos, son analíticamente cerradas. Aun así, diseña lo siguiente:

   a) Un algoritmo muestreador de Gibbs que actualice $\boldsymbol{\beta}$ y λ en dos bloques.

   b) Un algoritmo muestreador de Gibbs que actualice los componentes de $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$ y λ uno a la vez.