EST-46111: Fundamentos de Estadística

–Notas de Clase–

Juan Carlos Martínez-Ovando

ITAM

12 de diciembre de 2016

1. Estadística bayesiana no paramétrica

1.1. EBNP: Fundamentos

Preguntas comúnes

- ¿Cuántos grupos hay en un conjunto de datos? ¿Cuántas clases incluir en un modelo de mezclas? (Clasificación no supervisada)
- ¿Qué función especificar? (Predicción o clasificación supervisada)
- ¿Cuántos factores considerar empleando reducción de dimensionalidad? (Análisis de factores)

Concepción bayesiana

$$\begin{aligned} y|\theta &\sim F(y|\theta) \\ \theta &\sim \pi(\theta). \end{aligned}$$

Nota que $\pi(\theta)$ en realidad asigna probabilidad al modelo $F(\cdot|\theta)$. i.e.

$$\pi(\theta) \leftrightarrow \pi(F(\cdot|\theta)).$$

1.1. EBNP: Fundamentos

Preguntas comúnes

- ¿Cuántos grupos hay en un conjunto de datos? ¿Cuántas clases incluir en un modelo de mezclas? (Clasificación no supervisada)
- ¿Qué función especificar? (Predicción o clasificación supervisada)
- ¿Cuántos factores considerar empleando reducción de dimensionalidad? (Análisis de factores)

Concepción bayesiana

$$\begin{aligned} y|\theta &\sim F(y|\theta) \\ \theta &\sim \pi(\theta). \end{aligned}$$

Nota que $\pi(\theta)$ en realidad asigna probabilidad al modelo $F(\cdot|\theta)$. i.e.

$$\pi(\theta) \leftrightarrow \pi(F(\cdot|\theta)).$$

Concepción bayesiana no paramétrica (densidad)

Entonces, por qué no adoptar directamente lo siguiente

$$\begin{aligned} y|F &\sim F(y) \\ F &\sim \pi(F). \end{aligned}$$

Concepción bayesiana no paramétrica (regresión)

De igual forma, por qué no adoptar directamente lo siguiente

$$\begin{aligned} y|f &\sim \mathrm{N}(y|f, \lambda) \\ f &\sim \pi(f). \end{aligned}$$

1.1. EBNP: Fundamentos

Concepción bayesiana no paramétrica (densidad)

Entonces, por qué no adoptar directamente lo siguiente

$$\begin{aligned} y|F &\sim F(y) \\ F &\sim \pi(F). \end{aligned}$$

Concepción bayesiana no paramétrica (regresión)

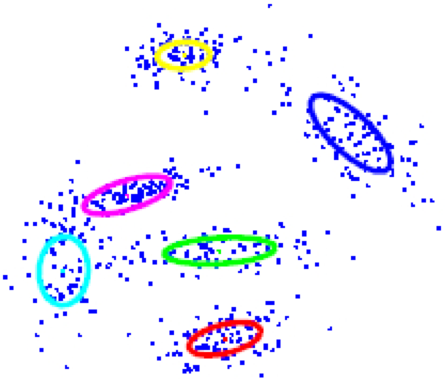
De igual forma, por qué no adoptar directamente lo siguiente

$$\begin{aligned} y|f &\sim \mathsf{N}(y|f,\lambda) \\ f &\sim \pi(f). \end{aligned}$$

1.2. EBNP: Mezcla de modelos

Clasificación no supervisada

Figura: Formación e identificación de clusters



1.2. EBNP: Mezcla de modelos

Proceso Dirichlet

Se define como una distribución sobre un espacio de distribuciones (i.e. medida de probabilidad sobre medidas de probabilidad)
Si G es la función de distribición sobre $\mathcal{X}$, tenemos

$$G \sim \mathsf{DP}(G|G_0, \alpha), \tag{1}$$

donde

- G_0 es una medida de probabilidad absolutamente continua
- α es un escalar positivo.

Comentarios

Para toda partición $(A_1, \ldots, A_n)$ de $\mathcal{X}$,

$$(G(A_1), \ldots, G(A_n)) \sim \mathsf{Dir}\left(\alpha G_0(A_1), \ldots, \alpha G_0(A_n)\right). \tag{2}$$

1.2. EBNP: Mezcla de modelos

Proceso Dirichlet

Se define como una distribución sobre un espacio de distribuciones (i.e. medida de probabilidad sobre medidas de probabilidad)
Si G es la función de distribición sobre $\mathcal{X}$, tenemos

$$G \sim \mathsf{DP}(G|G_0, \alpha), \tag{1}$$

donde

- G_0 es una medida de probabilidad absolutamente continua
- α es un escalar positivo.

Comentarios

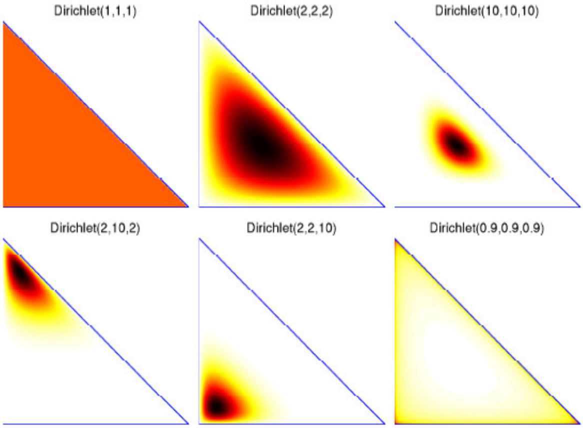
Para toda partición $(A_1, \ldots, A_n)$ de $\mathcal{X}$,

$$(G(A_1), \ldots, G(A_n)) \sim \mathsf{Dir}\left(\alpha G_0(A_1), \ldots, \alpha G_0(A_n)\right). \tag{2}$$

1.2. EBNP: Mezcla de modelos

Proceso Dirichlet

Figura: Asignación de distribuciones iniciales



1.2. EBNP: Mezcla de modelos

Proceso Dirichlet (forma)

Si G es un proceso Dirichlet, entonces,

$$G(\theta) = \sum_{i=1}^{\infty} \omega_i \delta_{\theta_i^*}(\theta), \tag{3}$$

donde

- $(\omega_i)_{i=1}^{\infty}$ se definen en el simpler de dimensión infinita
- $(\theta_i^*)_{i=1}^{\infty}$ son átomos aleatorios distintos definidos en Θ.

Se tiene que si θ_is son iid de G_0 entonces $\mathbb{E}(G) = G_0$.

Proceso Dirichlet (conjugacidad)

Si $G \sim \text{DP}(G|G_0, \alpha)$ y $\theta_1, \ldots, \theta_n$ son una m.a. de G, entonces

$$G|\theta_1, \ldots, \theta_n \sim \text{DP}\left(\frac{\alpha}{\alpha+n} G_0 + \frac{1}{\alpha+n} \sum_{j=1}^{n} \delta_{\theta_j}, \alpha+n\right). \tag{4}$$

1.2. EBNP: Mezcla de modelos

Proceso Dirichlet (forma)

Si G es un proceso Dirichlet, entonces,

$$G(\theta) = \sum_{i=1}^{\infty} \omega_i \delta_{\theta_i^*}(\theta), \tag{3}$$

donde

- $(\omega_i)_{i=1}^{\infty}$ se definen en el simpler de dimensión infinita
- $(\theta_i^*)_{i=1}^{\infty}$ son átomos aleatorios distintos definidos en Θ.

Se tiene que si θ_is son iid de G_0 entonces $\mathbb{E}(G) = G_0$.

Proceso Dirichlet (conjugacidad)

Si $G \sim \mathsf{DP}(G|G_0, \alpha)$ y $\theta_1, \ldots, \theta_n$ son una m.a. de G, entonces

$$G|\theta_1, \ldots, \theta_n \sim \mathsf{DP}\left(\frac{\alpha}{\alpha+n}G_0 + \frac{1}{\alpha+n}\sum_{j=1}^{n}\delta_{\theta_i}, \alpha+n\right). \tag{4}$$

Mezcla de proceso Dirichlet (forma)

Si $\mathcal{X}$ es elsoporte de una variable aleatoria observable y G es un proceso Dirichlet sobre un espacio parametral no observable θ, entonces,

$$
\begin{aligned}
P(x) &= \int_{\theta} \mathsf{N}(x|\theta) G(\mathrm{d}\theta) \\
&= \sum_{i=1}^{\infty} \omega_i \, \mathsf{N}(x|\theta_i^*) \\
&= \sum_{i=1}^{\infty} P(s=i|\boldsymbol{\omega}) \, \mathsf{N}(x|\theta_i^*, s=i),
\end{aligned}
$$

donde $\boldsymbol{s} = (s_1, \ldots, s_K, \ldots)$ dado $\boldsymbol{\omega}$ es multinomial, y las $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_K, \ldots)$ son Dirichlet (por lo que la conjugacidad se sigue).

Mezcla del proceso Dirichlet (clasificación)

Dada una muestra de datos, $x_1, \ldots, x_n$ se derivan probabilidades de clasificación

$$
P(s^{(j)} = i|\boldsymbol{s}_{-j}, \boldsymbol{\omega}) = \begin{cases} \frac{n_{-j,i}}{\alpha+(n-1)}, & \text{si i es representado} \\ \frac{\alpha}{\alpha+(n-1)}, & \text{si i no es representado} \end{cases} \tag{5}
$$

1.2. EBNP: Mezcla de modelos

Mezcla de proceso Dirichlet (forma)

Si $\mathcal{X}$ es elsoporte de una variable aleatoria observable y G es un proceso Dirichlet sobre un espacio parametral no observable θ, entonces,

$$
\begin{aligned}
P(x) &= \int_\theta \mathsf{N}(x|\theta)G(\mathrm{d}\theta) \\
&= \sum_{i=1}^{\infty} \omega_i \, \mathsf{N}(x|\theta_i^*) \\
&= \sum_{i=1}^{\infty} P(s=i|\boldsymbol{\omega}) \, \mathsf{N}(x|\theta_i^*, s=i),
\end{aligned}
$$

donde $\boldsymbol{s} = (s_1, \ldots, s_K, \ldots)$ dado $\boldsymbol{\omega}$ es multinomial, y las $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_K, \ldots)$ son Dirichlet (por lo que la conjugacidad se sigue).

Mezcla del proceso Dirichlet (clasificación)

Dada una muestra de datos, $x_1, \ldots, x_n$ se derivan probabilidades de clasificación

$$
P(s^{(j)} = i|\boldsymbol{s}_{-j}, \boldsymbol{\omega}) = \begin{cases} \frac{n_{-j,i}}{\alpha+(n-1)}, & \text{si i es representado} \\ \frac{\alpha}{\alpha+(n-1)}, & \text{si i no es representado} \end{cases} \tag{5}
$$

1.2. EBNP: Mezcla de modelos

Mezcla de proceso Dirichlet (inferencia)

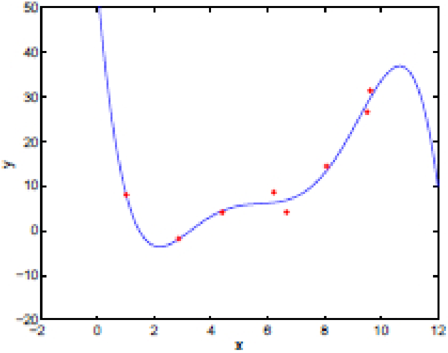
Se requieren métodos computacionales para estimar (inferir) y predecir en esta clase de modelos.

- *Gibbs sampling* (e.g. Escobar y West, 1995; Neal, 2000; Rasmussen, 2000)
- *Variational approximation* (Blei y Jordan, 2005)
- *Expectation propagation* (Minka y Ghahramani, 2003)
- *Hierarchical clustering* (Heller y Ghahramani, 2005)
- y muchos otros...

1.3. EBNP: Estimación de curvas

Suavizamiento

Figura: Curvas de respuesta



1.3. EBNP: Estimación de curvas

Proceso gaussiano

Supongamos que los datos toman valores en $\mathcal{X}$ y deseamos estimar una curva de respesta $f : \mathcal{X} \rightarrow \Re$, de tal forma que f es nuestro parámetro de interés.

Proceso gaussiano(definición)

Se dice que f es un proceso gaussiano si para todo $x_1, \ldots, x_n$ se sigue

$$P\left(f(x_1)\ldots, f(x_n)\right) = \mathsf{N}_n\left((f(x_1)\ldots, f(x_n))|(\mu(x_1)\ldots, \mu(x_n)), \mathbf{\Sigma}_n\right), \qquad (6)$$

donde

- μ es una función media y
- Σ es una función de covarianza.

1.3. EBNP: Estimación de curvas

Proceso gaussiano

Supongamos que los datos toman valores en $\mathcal{X}$ y deseamos estimar una curva de respesta $f : \mathcal{X} \to \Re$, de tal forma que f es nuestro parámetro de interés.

Proceso gaussiano(definición)

Se dice que f es un proceso gaussiano si para todo $x_1, \ldots, x_n$ se sigue

$$P\left(f(x_1)\ldots, f(x_n)\right) = \mathsf{N}_n\left((f(x_1)\ldots, f(x_n))|(\mu(x_1)\ldots, \mu(x_n)), \mathbf{\Sigma}_n\right), \qquad (6)$$

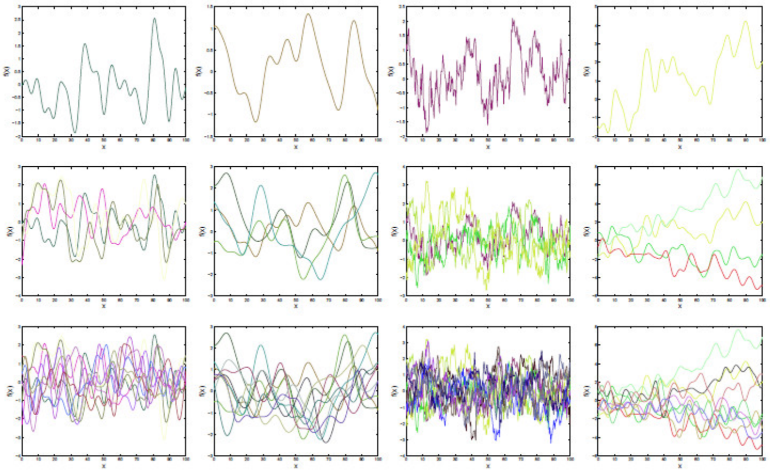
donde

- μ es una función media y
- Σ es una función de covarianza.

1.3. EBNP: Estimación de curvas

Proceso gaussiano

Figura: Procesos gaussianos con diferentes funciones de covarianza



1.3. EBNP: Estimación de curvas

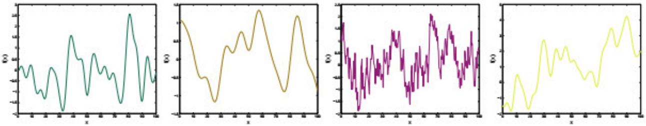
Proceso gaussiano (regresión)

Para los datos $\{(y_i, x_i)\}_{i=1}^n$, se puede definir

$$\begin{aligned} y_i | f, x_i &\sim \mathsf{N}(y_i | f(x_i), \lambda) \\ f &\sim \mathsf{GP}(f | 0, \boldsymbol{\Sigma}) \end{aligned}$$

Proceso gaussiano (muestras de la prior)

Figura: Simulación de distribuciones iniciales



1.3. EBNP: Estimación de curvas

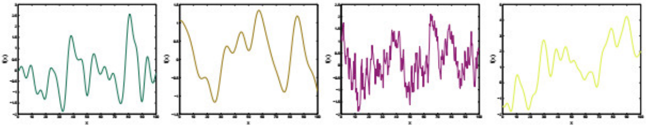
Proceso gaussiano (regresión)

Para los datos $\{(y_i, x_i)\}_{i=1}^{n}$, se puede definir

$$\begin{aligned} y_i|f, x_i &\sim \mathsf{N}(y_i|f(x_i), \lambda) \\ f &\sim \mathsf{GP}(f|0, \boldsymbol{\Sigma}) \end{aligned}$$

Proceso gaussiano (muestras de la prior)

Figura: Simulación de distribuciones iniciales



1.3. EBNP: Estimación de curvas

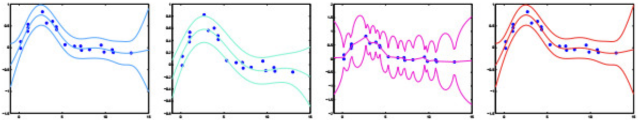
Proceso gaussiano (predicción)

Para los datos datos $= \{(y_i, x_i)\}_{i=1}^{n}$, la distribución predictiva es,

$$P(y_f | f, x_f, \text{datos}) \quad = \quad \int \mathsf{N}(y_f | f(x_f), \lambda)\, \mathsf{GP}(\mathrm{d}f | \text{datos})$$

Proceso gaussiano (muestras de la posterior)

Figura: Predicción de procesos gaussianos



1.3. EBNP: Estimación de curvas

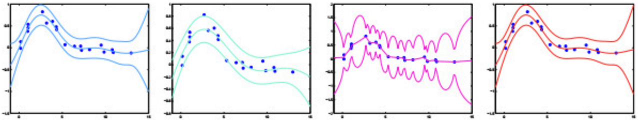
Proceso gaussiano (predicción)

Para los datos datos $= \{(y_i, x_i)\}_{i=1}^n$, la distribución predictiva es,

$$P(y_f | f, x_f, \text{datos}) = \int \mathsf{N}(y_f | f(x_f), \lambda)\, \mathsf{GP}(\mathrm{d}f | \text{datos})$$

Proceso gaussiano (muestras de la posterior)

Figura: Predicción de procesos gaussianos



1.3. EBNP: Estimación de curvas

Proceso gaussiano

Los procesos gaussianos se emplean en diferentes contextos

- Regresión (estimación)
- Regresión (predicción)
- Regresión (clasificación supervizada)
- Optimización bayesiana
- Series de tiempo

1.4. EBNP: Discusión

Discusión

- En ocasiones, se requieren distribuciones iniciales flexibles. Los modelos no paramétricos proporcionan una manera de definir modelos flexibles.
- Muchos modelos no paramétricos pueden derivarse partiendo de modelos paramétricos finitos y tomando el límite cuando el número de parámetros va al infinito.
- Los modelos no paramétricos pueden implementarse en la práctica con relativa facilidad.
- Hemos revisado los procesos gaussianos y los procesos de Dirichlet. Aún hay una gama extensa de modelos no paramétricos por explorar.

¡Gracias por su paciencia!

juan.martinez.ovando@itam.mx