Modelos de Regresión

– Enfoque Bayesiano General –

Juan Carlos Martínez-Ovando

ITAM - Ciencia de Datos

Fundamentos de Estadística
Maestría en Ciencia de Datos, ITAM
20 de octubre de 2015

Contents

Motivación

Estadística (bayesiana) no paramétrica

- Intenta reconocer ampliar la noción de incertidumbre en el modelo de manera más "flexible" que el enfoque paramétrico.
- Los parámetros no son los convencionales, en este caso *los parámetros son funciones (e.g., densidades, distribuciones, funciones de supervivencia, funciones de regresión, etc.).*

Problema de regresión

- Caracterizar la realización de una variable aleatoria y con un conjunto de covariables x

Motivación

Estadística (bayesiana) no paramétrica

- Intenta reconocer ampliar la noción de incertidumbre en el modelo de manera más "flexible" que el enfoque paramétrico.
- Los parámetros no son los convencionales, en este caso *los parámetros son funciones (e.g., densidades, distribuciones, funciones de supervivencia, funciones de regresión, etc.).*

Problema de regresión

- Caracterizar la realización de una variable aleatoria y con un conjunto de covariables $\boldsymbol{x}$

Motivación

Regresión bayesiana no paramétrica

- Integra el enfoque (bayesiano) no paramétrico al problema de regresión.
- En la literatura encontramos dos enfoques para esto:
    I. Englobar el problema de regresión convencional, con un enfoque no paramétrico.
    II. Estudiar el problema de regresión como un problema de probabilidades condicionales, estimando dicha probabilidad condicional no paramétricamente.

Estadística bayesiana no paramétrica

Caracterización

- Un modelo bayesiano no paramétrico es típicamente un modelo con un espacio parametral de dimensión infinita (numerable o denso).
- El espacio parametral se define como el espacio de todas las poibles soluciones de un proceso de aprendizaje.
- Por ejemplo, con variables aleatorias *iid*, $(Y_i)_{i\geq 1}$ con distribución $f(\cdot)$, el conjunto de posibles soluciones del proceso de aprendizaje son todas las distribuciones predictivas:

$$
\begin{aligned}
Y_1 &\sim f(y) \\
Y_2|y_1 &\sim f(y|y_1) \\
&\vdots \\
Y_{k+1}|y_1,\ldots,y_k &\sim f(y|y_1,\ldots,y_k) \\
&\vdots
\end{aligned}
$$

Aquí, $f_k(\cdot) = f(\cdot|y_1,\ldots,y_k)$ es desconocida (y aleatoria, bajo el enfoque bayesiano).

Estadística bayesiana no paramétrica

Apreciación subjetiva

- La incertidumbre sobre las $(f_k)_{k\geq 1}$ se manifiesta a través de una medida de probabilidad subjetiva, que liga la información observada con los eventos futuros inciertos.
- El **futuro**, Y_{k+1}, y **pasado**, $(y_1, \ldots, y_k)$ se conectan a través de una distribución Π sobre el espacio de todas las posibles realizaciones de f, i.e.

$$\mathcal{F} = \{f : \text{ tal que } f \text{ es una función de distribución}\}. \tag{1}$$

  i.e. f es positiva, monótona creciente, acotada en 1.
- El espacio $\mathcal{F}$ es "grande", puede incluir a todas las distribuciones Gaussianas, t-Snedecor, etc, todas en una clase.

Estadística bayesiana no paramétrica

Prodecimiento

- Pasado y futuro se ligan a través de la relación

$$f(y|y_1,\dots,y_k) \quad = \quad \int_{\mathcal{F}} f(y)\Pi(df|y_1,\dots,y_k), \tag{2}$$

donde

$$\Pi(df|y_1,\dots,y_k) \quad \propto \quad \prod_{i=1}^{k} f(y_i)\Pi(f)df. \tag{3}$$

Estadística bayesiana no paramétrica

Ejemplo: Modelos tipo mezclas

- Supongamos que f es caracterizado como

$$f(y) \quad = \quad \sum_{k=1}^{\infty} w_k f(y|\theta_k), \tag{4}$$

donde

  - $(w_k)_{k\geq 1}$ es una sucesión definida en el simplex de dimensión infinita.
  - $(\theta_k)_{k\geq 1}$ es una sucesión de parámetros en un espacio común Θ.
  - $f(\cdot|\theta)$ es una distribución paramétrica.

- En este caso, el espacio parametral $\mathcal{F}$ está definido como una biyección con producto cartesiano

$$\mathcal{F}_M = \otimes_{k=1}^{\infty} \left(\mathcal{W}_k \times \theta_k \right),$$

restringido a que $\otimes_{k=1}^{\infty} \mathcal{W}_k$ sea el simplex de dimensión infinito.

- Así, definir Π sobre $\mathcal{F}$ es equivalente a definir $\widetilde{\Pi}$ sobre $\mathcal{F}_M$.

Regresión bayesiana no paramétrica

Enfoque I

Se relaciona la varible de respuesta Y con un conjunto de covariables $\boldsymbol{x} = (x_1, \ldots, x_p)$ en la media,

$$\mathbb{E}(y|\boldsymbol{x}) = f(\boldsymbol{x}),$$

donde f es una función que mapea $\Re^p$ a $\Re$, i.e.

$$y = f(\boldsymbol{x}) + \varepsilon,$$

donde ε es *iid* G con $\mathbb{E}_G(\varepsilon) = 0$ y $var_G(\varepsilon) = \sigma^2$.

- En realidad, estamos suponiendo que Y dado $\boldsymbol{x}$ es tal que

  $$Y|\boldsymbol{x} \sim G(y|\boldsymbol{x}),$$

  tal que

  $$\mathbb{E}_G(y|\boldsymbol{x}) = f(\boldsymbol{x}) \quad \text{y} \quad var_G(y|\boldsymbol{x}) = \sigma^2.$$

- El espacio parametral es

  $$\mathcal{F} = \{f : \; f \text{ es una función de } \Re^p \text{ a } \Re\}.$$

Regresión bayesiana no paramétrica

Enfoque II

Se relaciona la varible de respuesta Y con un conjunto de covariables $\boldsymbol{x} = (x_1, \ldots, x_p)$ a través de la relación condicional,

$$Y|\boldsymbol{x} \sim F(y|\boldsymbol{x}).$$

- En este caso, todos los momentos condicionales de Y dado $\boldsymbol{x}$ son potencialmente función de $\boldsymbol{x}$.
- Sesgos, varianzas, asimetrías, etc., de Y pueden estar explicadas en función de $\boldsymbol{x}$.
- Más aun, pueden definirse diferentes grupos relacionales de y en $\boldsymbol{x}$ para diferentes configuracione de $\boldsymbol{x}$
- En este caso, el espacio parametral es

$$\mathcal{F} = \{f : f \text{ es una función de distribución condicional de } Y \text{ dado } \boldsymbol{x}\}.$$

Regresión: Enfoque I

Formulación

La regresión en medias, se define como

$$y_i = f(\boldsymbol{x_i}) + \varepsilon_i, \tag{5}$$

para $i = 1, , n$, donde

$$f(\boldsymbol{x}_i) = \sum_{j=1}^{\infty} \alpha_j \phi(\boldsymbol{x_j}), \tag{6}$$

donde

- $(\alpha_j)_{j\geq 1}$ es una sucesión de escalares, y
- $(\phi_j)_{j\geq 1}$ es una sucesión de funciones base de un espacio funcional (típicamente es el espacio L_1 de funciones integrables y medibles), tales que $\phi : \Re^p \to \Re$.
- $\mathbb{E}(\varepsilon_i) = 0$ y $var(\varepsilon_i) = \sigma^2$.

Regresión: Enfoque I

Alternativas de funciones base

- Descomposición espectral de funciones en L_1.
- Onduletas (*wavelets*)
- Funciones de base radial (*radial basis functions*).
- Kernel (en un enfoque tradicional).

En todos estos casos, un aspecto común es que

$$\phi_j(\boldsymbol{x}_i) = \phi(x_i; \theta_j), \tag{7}$$

donde $\phi(\cdot)$ es fija en j y θ_j es fija para todo j.
Por ejemplo, con bases radiales:

$$\phi_j(\boldsymbol{x}_i) = \phi(||\beta_j' x_i - \mu_j||), \tag{8}$$

donde $\phi(\cdot)$ es una función de $\Re$ en $\Re$ simétrica alrededor de 0.

Regresión: Enfoque I

Aspectos para su implementación

- Todos los modelos anteriores son esencialmente lineales en los parámetros, por lo que su implemnetación es relativamente simple.
- Sin embargo, en la práctica, decansan en truncamientos (aleatorios o arbitrarios) de la expansión de bases.

Regresión: Enfoque I

Linearización

La linearización del modelo consiste en:

- Truncar la expansión de f a

$$f(\boldsymbol{x_i}) \approx \sum_{j=1}^{K} \alpha_j \phi(\boldsymbol{x_j}), \tag{9}$$

  donde K puede ser fijo o aleatorio.

- Definir el modelo de regresión lineal en los parámetros $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_K)$, con la matriz de diseño:

$$\boldsymbol{D} = (\boldsymbol{d}_1, \ldots, \boldsymbol{d}_n), \tag{10}$$

  donde

$$\boldsymbol{d}_i = (\phi(\boldsymbol{x}_i; \theta_1), \ldots, \phi(\boldsymbol{x}_i; \theta_K)). \tag{11}$$

Regresión: Enfoque II

Formulación

Siguiendo la idea de mexclas de modelos, la distribución de regresión condicional puede definirse como:

$$f(y|\boldsymbol{x}) = \sum_{j=1}^{\infty} w_j(\boldsymbol{x}) N(y|\boldsymbol{\beta}_j' x, \sigma^2) \tag{12}$$

donde

- $(w_j(\boldsymbol{x}))_{j\geq 1}$ es una sucesión de pesos definidos en el simplex unidimensional, los cuales pueden estar en función de $\boldsymbol{x}$ (la inferencia es muy complicada en este caso).
- $N(y|\boldsymbol{\beta}_j' x, \sigma^2)$ es una distrubución Gaussiana convencional.

Regresión: Enfoque II

Racionalidad

- La racionalidad de este modelo es la de suponer que existe un número infinito de modelos de regresión caracterizados por diferentes coeficientes de regresón $(\beta_j)_{j\geq 1}$.
- Los datos se acoplan a las distintas configuraciones de regresión, con base en las actualizaciones de los pesos $(w_j)_{j\geq 1}$, o *develan* una nueva configuración particular.
- El uso de estos modelo trasciende al de regresión, ya que sirven para caracterizar **crluster** o agrupaciones en los datos.
- Se debe prestar atención a las condiciones de identificabilidad sobre $(\beta_j)_{j\geq 1}$.

Regresión: Enfoque II

Inferencia

Hace uso de métodos MCMC sofisticados, con:

- Pasos *de movimientos inter dimencionales* (e.g., transdimensional Gibbs samplers, reversible jump MCMC, etc.).
- Truncamiento estocásticos de sumas.
- Muestreadores aleatorios no conjugados (slice sampler, perfect samplers, etc.)

Algunas personas trabajando en el tema:

- Fuentes y Mena (UNAM)
- Walker (Texas)
- Zoubin Ghaharamani (Cambridge)
- Michael I. Jordan (Berkeley)
- Yee Whye Teh y Chris Holmes (Oxford)

Regresión: Enfoque II

Inferencia

Hace uso de métodos MCMC sofisticados, con:

- Pasos *de movimientos inter dimencionales* (e.g., transdimensional Gibbs samplers, reversible jump MCMC, etc.).
- Truncamiento estocásticos de sumas.
- Muestreadores aleatorios no conjugados (slice sampler, perfect samplers, etc.)

Algunas personas trabajando en el tema:

- Fuentes y Mena (UNAM)
- Walker (Texas)
- Zoubin Ghaharamani (Cambridge)
- Michael I. Jordan (Berkeley)
- Yee Whye Teh y Chris Holmes (Oxford)

Discusión

Comentarios

- El enfoque no paramétrico de regresión permite resolver problemas prácticos con bastante flexibilidad.
- La teoría descansa en nociones de espacios funcionales y procesos estocásticos.
- En la actualidad, estos problemas son aterrizados en la práctica.
- De hecho, muchos problemas teóricos on motivados por problemas prácticos.
- Se debe prestar atención a condiciones de convergencia asintótica.

Puntos de atención

- Decansan en métodos computacionales sofisticados (pero casi todos los modelos sofisticados lo hacen).
- Muchos de estos métodos ya han sido implementados y se distribuyen en software, principalmente gratuito.
- En el caso en que se deba desarrollar el software necesario, es bueno ya que tenemos control sobre los aspectos de implementación, sin descanar en *cajas negras*.

Discusión

Comentarios

- El enfoque no paramétrico de regresión permite resolver problemas prácticos con bastante flexibilidad.
- La teoría descansa en nociones de espacios funcionales y procesos estocásticos.
- En la actualidad, estos problemas son aterrizados en la práctica.
- De hecho, muchos problemas teóricos on motivados por problemas prácticos.
- Se debe prestar atención a condiciones de convergencia asintótica.

Puntos de atención

- Decansan en métodos computacionales sofisticados (pero casi todos los modelos sofisticados lo hacen).
- Muchos de estos métodos ya han sido implementados y se distribuyen en software, principalmente gratuito.
- En el caso en que se deba desarrollar el software necesario, es bueno ya que tenemos control sobre los aspectos de implementación, sin descanar en *cajas negras*.

Software

Paquetes y códigos

- `DPpackage` para R: Implementa una gran variedad de modelos bayesianos no paramétricos.
  URL: `cran.r-project.org/web/packages/DPpackage`
- `Hierarchical Bayesian compiler` implementa rutinas para modelos jerárquicos no paramétricos, implementados en Java.
  URL: `www.cs.utah.edu/ hal/HBC`
- `adaptor grammars` implementa modelos bayesianos composicionales no paramétricos.
  URL: `cog.brown.edu/ mj/Software`
- `MIT-Church project`, un proyecto wiki para modelos probabilísticos de cognición basado en rutinas MCMC.
  URL: `http://projects.csail.mit.edu/church/wiki/Church`
- `Bayesian variational bayes` rutinas en Matlab desarrolladas por Zoubin Ghaharamani y su grupo.
  URL: `mlg.eng.cam.ac.uk/zoubin/software.html`
- `Bayes Net toolbox` rutinas en Matlab desarrolladas por Kevin Murphy y su grupo.
  URL: `http://www.cs.ubc.ca/ murphyk/Software/`