Marco Bayesiano para el análisis de datos,
calibración de parámetros y modelamiento inverso

05

Modelamiento probabilístico

Universidad Industrial de Santander
U18 Fest

23–26 de septiembre de 2019

Teorema de Bayes

- **Objetivo**: Modelar datos
- Si tenemos observaciones y de un observable Y y un modelo de las observaciones parametrizado por Θ, podemos calcular la distribución de los parámetros dadas las observaciones usando

$$p(\theta \mid y) = \frac{p(y \mid \theta)p(\theta)}{p(y)} = \frac{p(y \mid \theta)p(\theta)}{\int p(y \mid \theta)p(\theta)\,\mathrm{d}\theta}$$

Nomenclatura

$$p(\theta \mid y) = \frac{p(y \mid \theta)p(\theta)}{p(y)} = \frac{p(y \mid \theta)p(\theta)}{\int p(y \mid \theta)p(\theta)\,\mathrm{d}\theta}$$

- $p(y \mid \theta)$: **Verosimilitud**
  El valor de la pdf de las observaciones dado el valor específico $\Theta = \theta$ de los parámetros. La verosimilitud indica qué tan posible es observar las (uh) observaciones dado θ

- $p(\theta)$: **(Distribución) Anterior**
  Codifica la información disponible o suposiciones acerca de la distribución de probabibilidad de Θ, i.e, indica qué valores de Θ son más o menos probables de acuerdo a la información *a priori*

Nomenclatura

$$p(\theta \mid y) = \frac{p(y \mid \theta)p(\theta)}{p(y)} = \frac{p(y \mid \theta)p(\theta)}{\int p(y \mid \theta)p(\theta)\,\mathrm{d}\theta}$$

- $p(\theta \mid y)$: **(Distribución) Posterior**
  pdf de Θ dadas las observaciones, i.e., indica qué valores de Θ son más o menos probables de acuerdo a las observaciones y la información *a priori*

- $p(y) = \int p(y \mid \theta)p(\theta)\,\mathrm{d}\theta$: **Verosimilitud marginal**
  Indica qué valores del observable Y son más o menos posibles dada la información *a priori* acerca de los parámetros del modelo

Nomenclatura

- Para calcular la distribución posterior sólo hace especificar las verosimilitud $p(y \mid \theta)$ y la distribución anterior, i.e, la distribución conjunta

$$p(y, \theta) = p(y \mid \theta)p(\theta)$$

- Ésta distribución conjunta se conoce como el *modelo probabilístico*

Tareas de regresión

- Queremos analizar la dependencia de un observable Y en una variable explanatoria x utilizando un modelo parametrizado por Θ
- Dadas las observaciones (X, y), podemos calcular la distribución de los parámetros:

$$p(\theta \mid y, X) = \frac{p(y \mid X, \theta)p(\theta)}{p(y \mid X)} = \frac{p(y \mid X, \theta)p(\theta)}{\int p(y \mid X, \theta)p(\theta)\,\mathrm{d}\theta}$$

donde

- $p(y \mid X, \theta)$: Verosimilitud
- $p(\theta)$: Anterior
- $p(\theta \mid y, X)$: Posterior

Regresión lineal ordinaria

- Tenemos un observable Y, que se asume contínuo y k variables explanatorias $x = (x_1, x_2, \ldots, x_3)$
- Un número n de observaciones de la forma

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_i \\ \vdots \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} x_{11} & x_{12} & \ldots & x_{1k} \\ x_{21} & x_{22} & \ldots & x_{2k} \\ \vdots & & & \vdots \\ x_{i1} & x_{i2} & \ldots & x_{ik} \\ \vdots & & & \vdots \\ x_{n1} & x_{n2} & \ldots & x_{nk} \end{bmatrix}$$

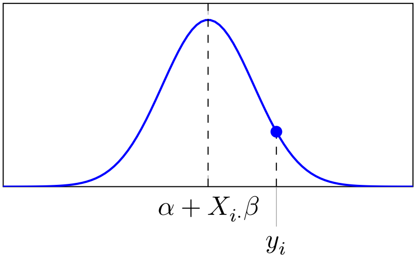
- Modelo lineal para la esperanza de las observaciones

$$\mathbb{E}(y_i \mid X, \theta) = \alpha + x_{i1}\beta_1 + x_{i2}\beta_2 + \ldots x_{ik}\beta_k$$
$$\mathbb{E}(y \mid X, \theta) = \alpha + X\beta$$

Verosimilitud normal

- Para modelar las observaciones, asumimos que cada observación es independiente, y modelada como una variable aleatoria normal con la esperanza dada por el modelo lineal y una cierta varianza

$$p(y_i \mid X, \theta) = \mathcal{N}(y \mid \alpha + X_{i\cdot}\beta, \sigma), \quad \theta = \{\alpha, \beta, \sigma\}$$



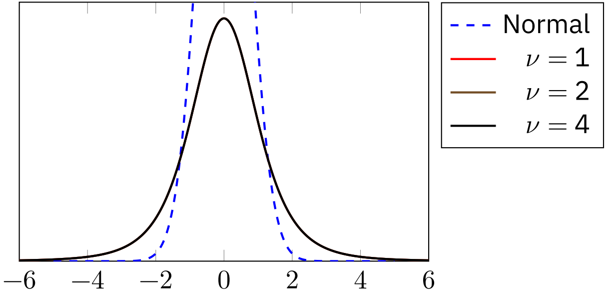
- La distribución conjunta de las observaciones es

$$p(y \mid X, \theta) = \prod_{i=1}^{n} p(y_i \mid X, \theta) = \mathcal{N}(y \mid \alpha + X\beta, \sigma^2 I)$$

Robustez contra valores extremos

- La distribución normal tiene colas "cortas". Por lo tanto, observaciones lejos del predictor $\alpha + X\beta$ para cierto valor de θ se consideran muy poco probables
- Éstas observaciones "arrastran" el estimador hacia ellas
- Para robustecer la inferencia contra valores extremos es necesario utilizar una verosimilitud con colas más largas

Robustez contra valores extremos



t-Distribución de Student

- Cómo decidir el valor de ν?
- Modelamiento jerárquico

Modelamiento jerárquico

- Las distribuciones anteriores de los parámetros, θ, se eligen usualmente como distribuciones parametrizadas de la forma $p(\theta \mid \nu)$ con sus propios parámetros ν, o *hiperparámetros*
- Podemos asignar *distribuciones hiperanteriores* $p(\nu)$ a éstos hiperparámetros

$$p(\theta, \nu \mid y, X) = \frac{p(y \mid X, \theta)p(\theta \mid \nu)p(\nu)}{\int p(y \mid X, \theta)p(\theta \mid \nu)p(\nu)\,\mathrm{d}\theta\mathrm{d}\nu}$$