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

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