Modelos Estadísticos para Clasificación y Regresión

Práctico 4 - Regresión Lineal Bayesiana

IMERL - FIng

13 de setiembre de 2023

FACULTAD DE INGENIERÍA
UDELAR

1. Repaso
    - Estimación Bayesiana
    - Gaussiana multivariada
    - Predicción
2. Práctico 4

1. Repaso
   - Estimación Bayesiana
   - Gaussiana multivariada
   - Predicción
2. Práctico 4

Estimación Bayesiana

Dado un conjunto de datos X, y y un modelo, consideraremos que los parámetros del modelo θ son variables aleatorias, con una **distribución *a priori*** $p(\theta)$.

Actualizaremos esta distribución con la información de los datos, de acuerdo a la regla de Bayes:

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}$$

Distribución a posteriori

A $p(\theta|y)$ se la conoce como **distribución *a posteriori***.

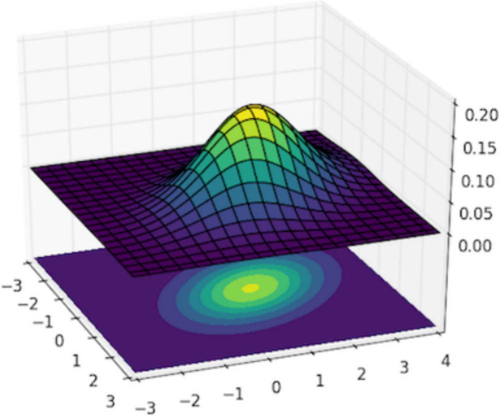
$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}$$

Observar que

- $p(y|\theta)$ es la verosimilitud.
- $p(y)$ es la verosimilitud marginal. Es decir, la marginal de la distribución conjunta $p(y,\theta)$, $p(y) = \int_0^1 p(y|\theta)p(\theta)d\theta$

Gaussiana multivariada

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{D/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\}$$



Regresión lineal bayesiana con gaussianas multivariadas

Asumimos que la verosimilitud $p(y|\theta)$ es una Gaussiana multivariada.

Para poder obtener una expresión exacta para la distribución a posteriori, precisamos elegir una distribución a priori $p(\theta)$ que sea conjugada de la verosimilitud gaussiana.

Nuestra distribución a priori será entonces una gaussiana multivariada con media μ_0 y matriz de covarianza Σ_0:

$$p(\theta|\mu_0, \Sigma_0) = \mathcal{N}(\mu_0, \Sigma_0)$$

Distribución a posteriori

Se puede probar que en estas condiciones la distribución a posteriori es

$$p(\theta|y, X, \sigma^2) = \mathcal{N}(\mu_\theta, \Sigma_\theta)$$

donde

$$\Sigma_\theta = (\tfrac{1}{\sigma^2} X^T X + \Sigma_0^{-1})^{-1}$$

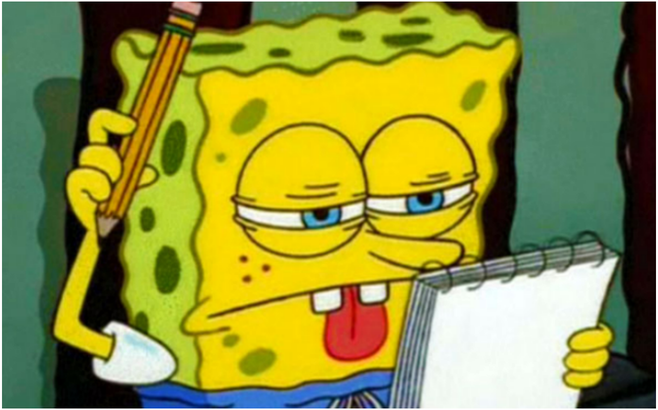
$$\mu_\theta = \Sigma_\theta(\tfrac{1}{\sigma^2} X^T y + \Sigma_0^{-1} \mu_0)$$

Predicción

Dado un nuevo dato x_{nuevo}, se tendrá una *distribución* de posibles valores de y_{nuevo}, dada por

$$p(y_{nuevo}|x_{nuevo}, X, y, \sigma^2) = \mathcal{N}(x_{nuevo}^T \mu_\theta, \sigma^2 + x_{nuevo}^T \Sigma_\theta x_{nuevo})$$

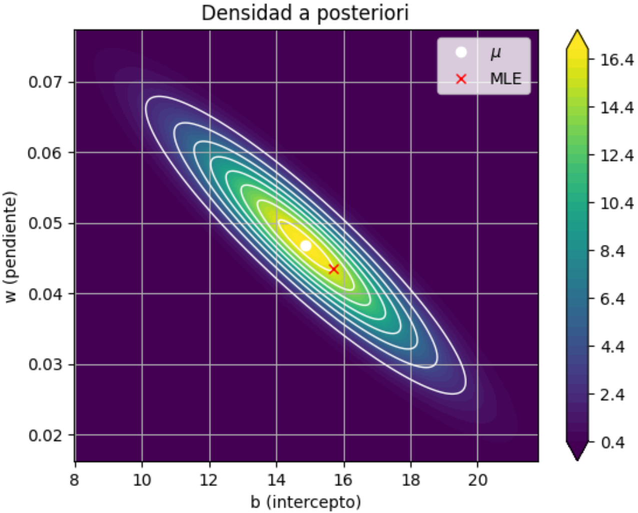
Preguntas?



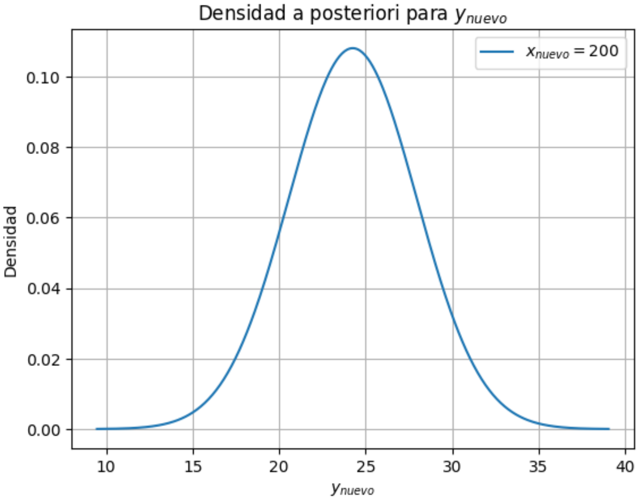
Práctico 4

```
MLE de b: 15.715950272533467
MLE de w: 0.04345049295663197
```

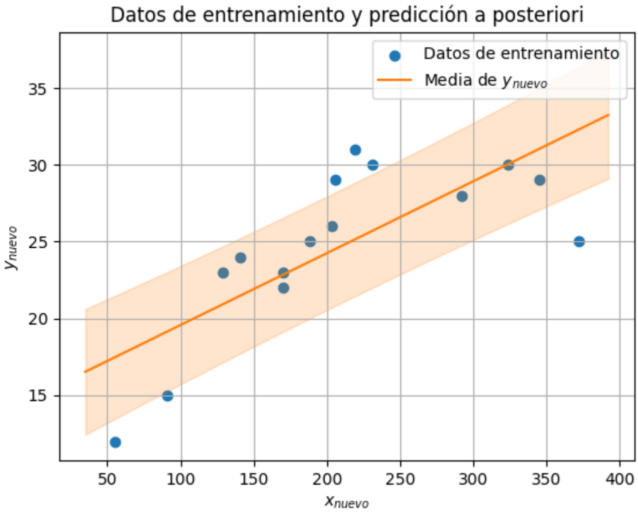
Práctico 4


Densidad a posteriori
μ
MLE
w (pendiente)
0.02
0.03
0.04
0.05
0.06
0.07
b (intercepto)
8
10
12
14
16
18
20
0.4
2.4
4.4
6.4
8.4
10.4
12.4
14.4
16.4

Práctico 4


Densidad a posteriori para y_{nuevo}
$x_{nuevo} = 200$
Densidad
0.00
0.02
0.04
0.06
0.08
0.10
10
15
20
25
30
35
40
y_{nuevo}

Práctico 4


Datos de entrenamiento y predicción a posteriori
Datos de entrenamiento
Media de y_{nuevo}
35
30
25
20
15
y_{nuevo}
50
100
150
200
250
300
350
400
x_{nuevo}