ITESO, Universidad Jesuita de Guadalajara

Maestría de Ciencia de Datos

Optimización Convexa

Tarea 2: Casos simples de maximum posteriori probability

Estudiante: Daniel Nuño
Profesor: Dr. Juan Diego Sanchez Torres
Fecha entrega: 21 de febrero de 2022

Introduction

Some problems in statistics and data science admit a representation through a probability distribution. For example, for the case of *unexplained* data $\{y_i\}_{i=1}^n$, possible models are $y_i \sim \text{iid}\mathcal{N}\left(\mu, \sigma^2\right)$ where $\mu \in \mathbb{R}$ is the mean and $\sigma^2 > 0$ is the variance, or for $y_i \in \{0, 1\}$, $y_i \sim \text{Bernoulli}(p)$ with $0 < p < 1$.

Similarly, for the case of input-output data $\{(x_i, y_i)\}_{i=1}^n$, a possible model is $y_i \sim \text{iid}\mathcal{N}\left(M(x_i; \theta), \sigma^2\right)$, where $M(x_i)$ is an adequate function with the parameters θ and $\sigma^2 > 0$.

Other problems involve previous information from the data, admitting a representation through a probability distribution and that previous knowledge. For example, for the case of *unexplained* data $\{y_i\}_{i=1}^n$, a possible model is $y_i \sim \text{iid}\mathcal{N}\left(\mu, \sigma^2\right)$ where $\mu \in \mathbb{R}$ is the mean and $\sigma^2 > 0$ is the variance and previous knowledge of the parameter μ has the model $\mu \sim \mathcal{N}\left(a, 1/\lambda^2\right)$.

When no previous information is available, the maximum likelihood approach offers a parametric estimation method. For the case of previous information modeled as a probability or a density function, the Bayes' theorem provides the parametric estimation process with the meaning of calculating a posterior distribution. Thus, the estimation is not longer about maximizing a likelihood but a posterior probability distribution. For this reason, this procedure receives the name of maximum a posteriori probability (MAP).

Activities

First, let start with a problem with no previous information, the *simple linear regression*.

Problem 1: Linear Regression Without Previous Information

Let the input-output data $\{(x_i, y_i)\}_{i=1}^n$. For this case, consider the model $y_i \sim \text{iid}\mathcal{N}\left(M(x_i; \theta), \sigma^2\right)$, with $\hat{y}_i = M(x_i) = mx_i + b$ and the parameters $\theta = [m, b]^T$. Use the ML approach to calculate estimators for the parameters θ and σ^2.

For this case, note that the estimators for $\theta = [m, b]^T$ coincide with those of the least squares procedure.

Problema 1

Problema 1: 21 Feb, 2022

Sea $p(Y|X) = N(Y|f(x), \sigma^2)$

$X \in \mathbb{R}$ son entradas y $Y \in \mathbb{R}$ son salidas con una relación entre $X - Y$ dada como $y = f(x) + \epsilon$ donde $\epsilon \sim N(0, \sigma^2)$

La función de Maxima Verosimilitud es $f(\epsilon_i)$ y buscamos el parametro $f(\epsilon_i) = y_i - \theta = \epsilon_i$.

$$f(\epsilon_i) = \frac{1}{\sigma\sqrt{2\pi}} e^{\left[\frac{-1}{2}\frac{(\epsilon_i - 0)^2}{\sigma^2}\right]}$$

$$L = \Pi \left(f = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{\left[\frac{-1}{2}\frac{\epsilon_i^2}{\sigma^2}\right]}\right.$$

$$\ln(L) = \ln L(\epsilon_i|\theta) = n\ln\sigma - \frac{n}{2}\ln 2\pi - \frac{1}{2\sigma^2}\sum_{i=1}^{n} \epsilon_i^2$$

$$\frac{\partial L}{\partial \theta} = -\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \theta)^2 = -\sum_{i=1}^{n}(y_i - \theta)^2$$

$$\frac{\partial L}{\partial \sigma^2} = \frac{-n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(y_i - \theta)^2$$

Problema 2

Now, consider a simple case of the MAP approach. Here, the prior distribution (which models the previously available information), and the likelihood function (that models the current information from the data) are normal.

Problem 2: A First Case of MAP

Let the output data $\{y_i\}_{i=1}^{n}$. For this case, consider the model $y_i \sim \text{iid}\mathcal{N}\left(M(x_i;\theta), \sigma^2\right)$, with $\hat{y}_i = M(x_i) = \mu$ and the parameter $\theta = \mu$. Besides, suppose previous knowledge of the parameter μ with the model $\mu \sim \mathcal{N}\left(a, 1/\lambda^2\right)$. Use the MAP approach to calculate the estimators for the parameters θ and σ^2.

Problema 2:

Partiendo de la distribución de probabilidad normal, sustituimos la media y la varianza.

$$f(\mu_i) = \frac{\lambda}{\sqrt{2\pi}} e^{\frac{-\lambda}{2}(\mu - a)^2}$$

la probabilidad donde $\theta = \mu$

$$f(\theta | y_i) = \frac{f(y_i | \theta) \cdot f(\theta)}{f(y_i)}$$

21 feb, 20[illegible]

$$\ln(f(\theta|y_i)) = \ln L(f(y_i|\theta)) + \ln L(f(\theta)) - \ln(f(y_i))$$

$$\ln L(f(y_i|\theta)) = n \ln \sigma - \frac{n}{2}\ln 2\pi - \frac{1}{\sigma^2}\sum (y_i - \mu)^2$$

$$\ln L(f\theta) = \ln\left(\frac{\lambda}{\sqrt{2\pi}}\right) - \frac{\lambda^2}{2}(\mu - a)^2$$

$$\ln L(f(y_i)) = \ln L(f(y_i))$$

para θ:

$$\frac{\partial L(\theta|y_i)}{\partial \theta} \quad \ln L(f(y_i|\theta)) + \ln(\ln f(\theta)) = 0$$

$$\frac{1}{\sigma^2}\sum (y_i - \theta) - \lambda^2(\theta - a) = 0$$

$$\theta = \frac{\sum y_i + \lambda^2 a}{n + \lambda^2\sigma^2}$$

para σ^2

$$\frac{\partial L(L(\theta|y_i))}{\partial \sigma^2} \quad \left(n \ln \sigma - \frac{1}{2\sigma^2}\sum (y_i - \theta)\right) = 0$$

$$\frac{-n}{\sigma^2} + \frac{1}{2\sigma^4}\sum^{n} (y_i - \theta)^2 = 0$$

Problema 3

The following case considers another application of the MAP approach. Here, the prior distribution (which models the previously available information) and the likelihood function (that models the current information from the data) are normal. Furthermore, there is an input variable x.

Problem 3: Linear Regression With Previous Information

Let the input-output data $\{(x_i, y_i)\}_{i=1}^{n}$. For this case, consider the model $y_i \sim \text{iid}\mathcal{N}\left(M(x_i;\theta), \sigma^2\right)$, with $\hat{y}_i = M(x_i) = mx_i + b$ and the parameters $\theta = [m, b]^T$. Besides, suppose previous knowledge of the parameter m with the model $m \sim \mathcal{N}\left(a, 1/\lambda^2\right)$. Use the MAP approach to calculate the estimators for the parameters θ and σ^2.

Problema 3: 21. Feb, 2022

Usando el teorema de Bayes

$$\ln L(f(\theta|\varepsilon_i)) = \ln L(f(\varepsilon_i|\theta)) + \ln L(f(\theta)) - \ln(f(\varepsilon_i))$$

donde

$$\varepsilon_i = y_i - mx_i - b$$

$$\theta = \mu$$

$$\ln L(f(\varepsilon_i|\theta)) = n\ln\sigma - \frac{n}{2}\ln 2\pi - \frac{1}{2\sigma^2}\sum(\varepsilon_i - \mu)^2$$

$$\ln L(f(\theta)) = \ln\left(\frac{\lambda}{\sqrt{2\pi}}\right) - \frac{\lambda^2}{2}(\mu - a)^2$$

$$\ln L(P(\varepsilon_i)) = \ln L(p(\varepsilon_i))$$

Para obtener θ:

$$\frac{\partial L(\theta|\varepsilon_i)}{\partial\theta} = \frac{1}{\sigma^2}\sum(\varepsilon_i - \theta) - \lambda^2(\theta - a) = 0$$

$$\theta = \frac{\sum\varepsilon_i + \lambda^2 a}{n + \lambda^2\sigma^2}$$

para obtener θ:

$$\frac{\partial L}{\partial\sigma^2} = \frac{-n}{\sigma^2} + \frac{1}{2\sigma^4}\sum(\varepsilon_i - \theta) = 0$$

$$\frac{n}{\sigma^2} = \frac{1}{2\sigma^4}\sum(\varepsilon_i - \theta)$$

$$\frac{2\sigma^4}{\sigma^2} = n\sum(\varepsilon_i - \theta)$$

$$\sigma = \left(\frac{n}{2}\sum(\varepsilon_i - \theta)\right)^{1/2}$$

Problema 4

Finally, the following case considers a binary response variable. This example helps to understand logistic regression, an essential model for classification purposes.

Problem 4: The Bernoulli distribution and MAP

Consider the Bernoulli distribution $p(x \mid \mu) = \mu^x(1-\mu)^{1-x}$ with $x \in \{0,1\}$.

1. Show that $p(x \mid \mu)$ can be written of the form $p(x \mid \mu) = \exp\left[x \ln \frac{\mu}{1-\mu} + \ln(1-\mu)\right]$.

2. Let $\mu \sim \text{Beta}(\alpha, \beta)$, so μ has the probability density function $p(\mu \mid \alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\mu^{\alpha-1}(1-\mu)^{\beta-1}$ where $\Gamma(\cdot)$ is the Gamma function. Let $x \in \{0,1\}$ be distributed according to the Bernoulli distribution with parameter $\theta \in [0,1]$, that is, $p(x = 1 \mid \theta) = \theta$. This can also be expressed as $p(x \mid \theta) = \theta^x(1-\theta)^{1-x}$. Let θ be distributed according to a Beta distribution with parameters α, β, that is, $p(\theta \mid \alpha, \beta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}$.

   Show that multiplying the Beta and the Bernoulli distributions, it follows $p(\theta \mid x, \alpha, \beta) = p(x \mid \theta)p(\theta \mid \alpha, \beta) \propto p(\theta \mid \alpha + x, \beta + (1-x))$

3. Let the data-set $D = \{y_i\}_{i=1}^n$ such that $y_i \in \{0,1\}$. Propose an adequate model a its respective optimal parameter estimation procedure for the case of the not availability of previous information.

4. Let the data-set $D = \{y_i\}_{i=1}^n$ such that $y_i \in \{0,1\}$. Propose an adequate model a its respective optimal parameter estimation procedure for the cases of the availability of previous information.

Problema 4 21 feb, 2022

4.1 queremos que $p(x|\mu) = \mu^x (1-\mu)^{1-x}$ se convierta en $e(x \ln(\frac{\mu}{1-\mu}) + \ln(1-\mu))$

$$\ln(p(x|\mu)) = x \ln \mu + (1-x) \ln(1-\mu)$$
$$= \underline{x \ln \mu} + \ln(1-\mu) - \underline{x \ln(1-\mu)}$$
$$= x \ln\left(\frac{\mu}{1-\mu}\right) + \ln(1-\mu)$$
$$p(x|\mu) = e^{\left(x \ln\left(\frac{\mu}{1-\mu}\right) + \ln(1-\mu)\right)}$$

4.2 $\mu \sim \text{Beta}(\alpha, \beta)$, $p(\Gamma|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$

$$p(x|\theta) = \theta^x (1-\theta)^{1-x}$$
$$p(\theta|\alpha,\beta) \propto \theta^{\alpha-1} (1-\theta)^{\beta-1}$$
$$p(\theta|x,\alpha,\beta) = p(x|\theta)\, p(\theta|\alpha,\beta) \propto p(\theta|\alpha+x, \beta(1-x))$$
$$p(x|\theta)\, p(\theta|\alpha,\beta) = \left[\theta^x (1-\theta)^{1-x}\right]\left[\theta^{\alpha-1} (1-\theta)^{\beta-1}\right]$$
$$= \theta^{x+\alpha-1} (1-\theta)^{(1-x)+(\beta-1)}$$
$$= \theta^{(\alpha+x)-1} (1-\theta)^{[\beta+(1-x)]-1}$$
$$p(\theta|x,\alpha,\beta) \propto p(\theta|\alpha+x, \beta+(1-x))$$

4.3 $L(D,\theta) = \prod p^{y_i}(1-p)^{1-y_i} = p^{\sum y_i}(1-p)^{n-\sum y_i}$ 21 Feb, 2022

donde $\theta = p$

$$\ln(L(D|\theta)) = \sum y_i \ln p + (n - \sum_i y_i)\ln(1-p)$$

$$\frac{\partial \ln(L(D|\theta))}{\partial p} = \frac{\sum y_i}{p} + \frac{(n-\sum y_i)(-1)}{1-p} = 0$$

$$= \frac{(1-p)\sum y_i - p(n-\sum y_i)}{p(1-p)} = 0$$

$$= \frac{\sum y_i - pn}{p(1-p)} = 0$$

por lo tanto $p = \frac{\sum y_i}{n} = \bar{y}$

4.4 sea f $p \sim N(0, \sigma^2)$

$$f(p) = \frac{1}{\sigma\sqrt{2\pi}} e^{\left[-\frac{1}{2}\frac{(p-0)^2}{\sigma^2}\right]}$$

entonces MAP

$$\ln(f(p|D)) = \ln(f(D|p)) + \ln(f(p)) - \ln(f(D))$$

$$\ln(f(D|p)) = \sum y_i \ln p + (n - \sum y_i)\ln(1-p)$$

$$\ln(f(p)) = n\ln(\sigma^2) - \frac{n}{2}\ln 2\pi - \frac{1}{2\sigma^2}\sum p^2$$

$$\ln(f(D)) = \ln(f(D))$$

21 feb, 2022

$$\frac{\partial f(p|D)}{\partial p} = \frac{\sum y_i}{p} - \frac{(n - \sum y_i)}{1-p} - \frac{1}{\sigma^2} \sum (p) = 0$$

$$= \frac{\sum y_i}{p} - \frac{n \sum y_i}{1-p} - \frac{1}{\sigma^2} np = 0$$

$$= \frac{\sum y_i (1-p) \sigma^2 - (n - \sum y_i) p \sigma^2 - n(p^3 - p^2)}{\sigma^2 p (1-p)} = 0$$

$$\frac{\partial f(p|D)}{\partial p} = \sum y_i = \frac{-n(p^3 - p^2 + p\sigma^2)}{(1-p)\sigma^2 + p\sigma^2}$$