Problem Set 4 - PGM

Tassos Karageorgiadis

aid21002@uom.edu.gr
tkarageorgiadis@gmail.com

November 2020

1 Problem 2

The Gaussian distribution is given from the following type

$$\mathbf{f}(\mathbf{x}) = \frac{1}{\sqrt{2 \cdot \pi \cdot \sigma^2}} \cdot \exp^{-\dfrac{(x-\mu)^2}{2\sigma^2}} \tag{1}$$

The Maximum Likelihood Estimation (log-likelihood) for the variance $\hat{\sigma}^2$, is

$$
\begin{aligned}
\mathcal{LL}(\hat{\sigma}^{\in}) &= \ln \prod_{n=1}^{N} p(x_k \mid \sigma^2) \\
&= \ln \prod_{n=1}^{N} \frac{1}{\sqrt{2 \cdot \pi \cdot \sigma^2}} \cdot \exp^{-\dfrac{(x-\mu)^2}{2\sigma^2}} \\
&= \sum_{n=1}^{N} \Big(\ln(\frac{1}{\sqrt{2 \cdot \pi \cdot \sigma^2}}) + \ln\big(\exp^{-\dfrac{(x-\mu)^2}{2\sigma^2}} \big) \Big) \\
&= \sum_{n=1}^{N} \left(\ln 1 - ln\sqrt{2\pi\sigma^2} + \Big(\frac{-(x_n-\mu)^2}{2\sigma^2} \cdot \ln e \Big) \right) \\
&= \sum_{n=1}^{N} \left(- ln\sqrt{2\pi\sigma^2} + \frac{-(x_n-\mu)^2}{2\sigma^2} \right) \\
&= \sum_{n=1}^{N} \left(- \frac{ln 2\pi\sigma^2}{2} - \frac{(x_n-\mu)^2}{2\sigma^2} \right) \\
&= \sum_{n=1}^{N} \left(- \frac{ln 2\pi\sigma^2}{2} \right) + \sum_{n=1}^{N} \left(- \frac{(x_n-\mu)^2}{2\sigma^2} \right) \\
&= - \frac{N \cdot ln 2\pi\sigma^2}{2} - \frac{1}{2\sigma_2} \sum_{n=1}^{N} (x_n-\mu)^2
\end{aligned}
\tag{2}
$$

$$
\boxed{\mathcal{LL}() = - \frac{N \cdot ln 2\pi\sigma^2}{2} - \frac{1}{2\sigma_2} \sum_{n=1}^{N} (x_n-\mu)^2}
\tag{3}
$$

In order to maximize the derivative of LL must be equal to 0,

$$
\begin{aligned}
\frac{\partial \mathcal{LL}}{\partial \sigma^2} &= \frac{\partial}{\partial \sigma^2} \left(\frac{-N \cdot ln 2\pi\sigma^2}{2} - \frac{1}{2\sigma_2} \sum_{n=1}^{N} (x_n-\mu)^2 \right) = 0 \\
&= - \frac{N}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{n=1}^{N} (x_n-\mu)^2 = 0 \\
&\Leftrightarrow \frac{1}{2\sigma^4} \sum_{n=1}^{N} (x_n-\mu)^2 = \frac{N}{2\sigma^2} \text{ ,multiply with } 2\sigma^2 \\
&\Leftrightarrow \frac{1}{\sigma^2} \sum_{n=1}^{N} (x_n-\mu)^2 = N \\
&\Leftrightarrow \sigma^2 = \frac{1}{N} \sum_{n=1}^{N} (x_n-\mu)^2
\end{aligned}
\tag{4}
$$

$$\boxed{\hat{\sigma}^2_{ML} = \frac{1}{N}\sum_{n=1}^{N}(x_n - \mu)^2} \tag{5}$$

From (3) if we get the derivative for μ,

$$\begin{aligned}
\frac{\partial \mathcal{L}\mathcal{L}}{\partial \mu} &= \frac{\partial}{\partial \mu}\left(\frac{-N \cdot ln 2\pi\sigma^2}{2} - \frac{1}{2\sigma_2}\sum_{n=1}^{N}(x_n - \mu)^2\right) \\
&= \frac{\partial}{\partial \mu}\left(\frac{-N \cdot ln 2\pi\sigma^2}{2}\right) + \sum_{n=1}^{N}\frac{\partial}{\partial \mu}\left(-\frac{1}{2\sigma_2}(x_n - \mu)^2\right) \\
&= 0 + \sum_{n=1}^{N}\left(\cancel{\frac{\partial}{\partial \mu}\left(-\frac{1}{2\sigma_2}\right)\left((x_n - \mu)^2\right)} + \left(\frac{1}{2\sigma_2}\right)\frac{\partial}{\partial \mu}\left((x_n - \mu)^2\right)\right) \\
&= \frac{-1}{2\sigma^2}\sum_{n=1}^{N}\frac{\partial}{\partial}(x_n - \mu)^2 \\
&= \frac{-1}{2\sigma^2}\sum_{n=1}^{N} 2 \cdot x_n - \mu \cdot (-1) \\
&= \frac{1}{\sigma^2}\sum_{n=1}^{N}(x_n - \mu)
\end{aligned} \tag{6}$$

Then the derivative must be equal to zero to find maximum,

$$\begin{aligned}
&\frac{1}{\sigma^2}\sum_{n=1}^{N}(x_n - \mu) = 0 \\
&\Leftrightarrow \sum_{n=1}^{N}(x_n - \mu) = 0 \\
&\Leftrightarrow \sum_{n=1}^{N} x_n - \sum_{n=1}^{N}\mu = 0 \\
&\Leftrightarrow \sum_{n=1}^{N} x_n - N \cdot \mu = 0 \\
&\Leftrightarrow \mu = \frac{1}{n}\sum_{n=1}^{N} x_n
\end{aligned} \tag{7}$$

$$\boxed{\hat{\mu}_{ML} = \frac{1}{n}\sum_{n=1}^{N} x_n} \tag{8}$$

2 Problem 3

With the equations that proved above in the problem 2, using maximum-likelihood estimation we can calculate the μ, and σ^2 for the two classes. The *class1* has a mean value of $\mu_{class1} = 0.26$, and standard deviation of $\sigma^2_{class1} = 0.1287$, so the Gaussian distribution is given by the

$$\mathbf{f(x_1)} = \frac{1}{\sqrt{2 \cdot \pi \cdot 0.1287}} \cdot \exp^{-\frac{(x_1 - 0.26)^2}{2 \cdot 0.1287}} \tag{9}$$

as for the *class2* the mean value is $\mu_{class2} = 0.863$, and the standard deviation is $\sigma^2_{class2} = 0.1108$,

$$\mathbf{f(x_2)} = \frac{1}{\sqrt{2 \cdot \pi \cdot 0.1108}} \cdot \exp^{-\frac{(x_2 - 0.863)^2}{2 \cdot 0.1108}} \tag{10}$$

We can see the distribution for each class in the following figure,

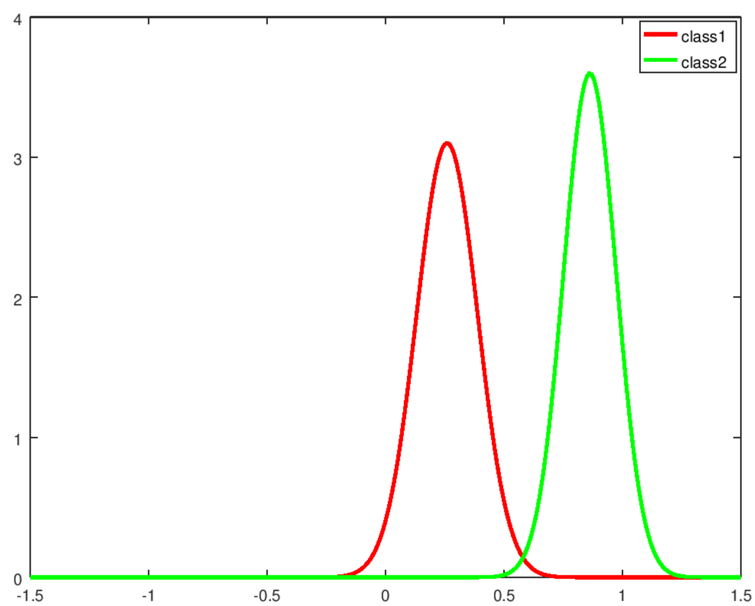

Figure 1: $Gaussian Distributions for class 1, class 2$

To calculate the probability of the event $x = 0.6$ to belong to class 1 we used Octave code again,

```
p1_06 = (1/sqrt(2*pi*std1) )*exp(-((0.6-mean1)^2)/(2*std1))
```

where mean1 is equal to $\mu_{class1} = 0.26$, and std1 is equal to $\sigma^2_{class1} = 0.1287$, so the result,

$$\boxed{p(x_1 = 0.6) = 0.70971} \tag{11}$$

.

3 Problem 4

For ML(Maximum Likelihood) we know that is asymptotically unbiased, so $\lim_{x\to\infty} E[\hat{(\theta)}_{ML}] = \theta_0$, we have to prove something similar for CL estimator?!

$$CL(\theta) = \frac{1}{N}\sum_{n=1}^{N} \log p(y_n \mid x_n, \theta) \tag{12}$$

$$CL(\theta)_{N\longrightarrow\infty} = \iint_0^\infty \log p(y_n x_n, \mid \theta)\, dx\, dy \tag{13}$$

4 Problem 5

$$B(x \mid \alpha, \beta) = \frac{1}{B(\alpha,\beta)} x^{\alpha-1} \cdot (1-x)^{\beta-1} \tag{14}$$

$$
\begin{aligned}
\mathcal{LL}(\alpha,\beta) &= \log \prod_{n=1}^{N} p(x_n \mid \alpha,\beta) \\
&= \log \prod_{n=1}^{N} \frac{1}{B(\alpha,\beta)} x_n^{\alpha-1} \cdot (1-x_n)^{\beta-1} \\
&= \sum_{n=1}^{N} \log\left(\frac{1}{B(\alpha,\beta)} x_n^{\alpha-1} \cdot (1-x_n)^{\beta-1}\right) \\
&= \sum_{n=1}^{N} \Bigg(\log(x_n^{\alpha-1}) - \log(B(\alpha,\beta)) + log((1-x_n)^{\beta-1}) \Bigg) \\
&= \sum_{n=1}^{N} \Bigg((\alpha-1)\cdot\log(x_n) - \log(B(\alpha,\beta)) + (\beta-1)\cdot\log(1-x_n) \Bigg) \\
&= (\alpha-1)\cdot\sum_{n=1}^{N}\log(x_n) - \sum_{n=1}^{N}\log(B(\alpha,\beta)) + (\beta-1)\sum_{n=1}^{N}\log(1-x_n) \\
&= (\alpha-1)\cdot\sum_{n=1}^{N}\log(x_n) - N\cdot\log(B(\alpha,\beta)) + (\beta-1)\sum_{n=1}^{N}\log(1-x_n)
\end{aligned} \tag{15}
$$

We know that

$$
\boxed{B(\alpha,\beta) = \frac{\Gamma(\alpha)\cdot\Gamma(\beta)}{\Gamma(\alpha+\beta)}} \tag{16}
$$

$$
\boxed{\psi(x) = \frac{\partial \log(\Gamma(x))}{\partial x}} \tag{17}
$$

From (15) we can calculate the partial derivative for α, β so,

$$
\begin{aligned}
\frac{\partial \mathcal{LL}}{\partial \alpha} &= \frac{\partial}{\partial \alpha}\Bigg((\alpha-1)\cdot\sum_{n=1}^{N}\log(x_n) - N\cdot\log(B(\alpha,\beta)) + (\beta-1)\sum_{n=1}^{N}\log(1-x_n)\Bigg) \\
&= \Bigg(\frac{\partial}{\partial \alpha}(\alpha-1)\cdot\sum_{n=1}^{N}\log(x_n) - \frac{\partial}{\partial \alpha}N\cdot\log(B(\alpha,\beta)) + 0\Bigg) \\
&= \sum_{n=1}^{N}\log(x_n)\cdot\cancel{\frac{\partial}{\partial \alpha}(\alpha-1)}^{\mathbf{1}} - N\frac{\partial}{\partial \alpha}\cdot\log(B(\alpha,\beta)) \;\textit{,from (16),} \\
&= \sum_{n=1}^{N}\log(x_n) - N\frac{\partial}{\partial \alpha}\cdot\log\left(\frac{\Gamma(\alpha)\cdot\Gamma(\beta)}{\Gamma(\alpha+\beta)}\right) \;\textit{,using}\; \Big[\log(\frac{x}{y}) = \log(x)-\log(y)\Big] \\
&= \sum_{n=1}^{N}\log(x_n) - N\cdot\psi(\alpha) + N\cdot\psi(\alpha+\beta)
\end{aligned} \tag{18}
$$

Using the above equation,

$$
\begin{aligned}
\frac{\partial \mathcal{LL}}{\partial \beta} &= \Bigg(0 - \frac{\partial}{\partial \beta}N\cdot B(\alpha,\beta) + \frac{\partial}{\partial \beta}(\beta-1)\sum_{n=1}^{N}\log(1-x_n)\Bigg) \\
&= -N\frac{\partial}{\partial \beta}\cdot\log B(\alpha,\beta) + \cancel{\frac{\partial}{\partial \beta}(\beta-1)}^{\mathbf{1}}\frac{\partial}{\partial \beta}\sum_{n=1}^{N}\log(1-x_n) \\
&= -N\cdot\frac{\partial}{\partial \beta}\log\left(\frac{\Gamma(\alpha)\cdot\Gamma(\beta)}{\Gamma(\alpha+\beta)}\right) + \sum_{n=1}^{N}\log(1-x_n) \\
&= \sum_{n=1}^{N}\log(1-x_n) - N\cdot\psi(\beta) + N\cdot\psi(\alpha+\beta)
\end{aligned} \tag{19}
$$

We can use ML to find parameters α,β by finding were the partial derivative for each one, is equal to zero.