Andres Fernandez – 5692442 – fr_andres@msn.com

Exercise 1

For the one-dimensional case, show that the product of two normal distributions with means μ_1, μ_2 and variances σ_1^2, σ_2^2 is proportional to a normal distribution with mean between the original two means and variance smaller than either of the original variances.

i)

Again, the one-dimensional PDF of the normal distribution is as follows:

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right) \tag{1}$$

ii)

With this, the product of two normal PDFs is:

$$\begin{aligned}
&\frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{1}{2}\frac{(x-\mu_1)^2}{\sigma_1^2}\right) \cdot \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left(-\frac{1}{2}\frac{(x-\mu_2)^2}{\sigma_2^2}\right) \\
&= \frac{1}{\sqrt{2\pi\sigma_1^2}} \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left(-\frac{1}{2}\frac{(x-\mu_1)^2}{\sigma_1^2}\right) \exp\left(-\frac{1}{2}\frac{(x-\mu_2)^2}{\sigma_2^2}\right) \\
&= \frac{1}{2\pi\sigma_1\sigma_2} \exp\left(-\frac{1}{2}\left(\frac{(x-\mu_1)^2}{\sigma_1^2} + \frac{(x-\mu_2)^2}{\sigma_2^2}\right)\right) \\
&= \frac{1}{2\pi\sigma_1\sigma_2} \exp\left(-\frac{1}{2}\left(\frac{\sigma_2^2(x-\mu_1)^2}{\sigma_1^2\sigma_2^2} + \frac{\sigma_1^2(x-\mu_2)^2}{\sigma_1^2\sigma_2^2}\right)\right) \\
&= \frac{1}{2\pi\sigma_1\sigma_2} \exp\left(-\frac{1}{2}\frac{\sigma_2^2(x-\mu_1)^2 + \sigma_1^2(x-\mu_2)^2}{\sigma_1^2\sigma_2^2}\right)
\end{aligned} \tag{2}$$

iii)

Expanding the second term in the exponent (see attached paper; *"P. A. Bromiley: Products and Convolutions of Gaussian Probability Density Functions"*), follows:

$$\begin{aligned} &\frac{\sigma_2^2(x-\mu_1)^2+\sigma_1^2(x-\mu_2)^2}{\sigma_1^2\sigma_2^2} \\ \text{(expand/reformulate)} &= \frac{(\sigma_1^2+\sigma_2^2)x^2-2(\mu_1\sigma_2^2+\mu_2\sigma_1^2)x+\mu_1^2\sigma_2^2+\mu_2^2\sigma_1^2}{\sigma_1^2\sigma_2^2} \\ \text{(div. through coeff. of } x^2) &= \frac{x^2-2(\frac{\mu_1\sigma_2^2+\mu_2\sigma_1^2}{\sigma_1^2+\sigma_2^2})x+\frac{\mu_1^2\sigma_2^2+\mu_2^2\sigma_1^2}{\sigma_1^2+\sigma_2^2}}{\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}} \\ \text{(compress quad.)} &= \frac{(x-\frac{\mu_1\sigma_2^2+\mu_2\sigma_1^2}{\sigma_1^2+\sigma_2^2})^2}{\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}} \end{aligned} \tag{3}$$

Which is analogous to the expression in the exponent of the normal, one-dimensional PDF $\frac{(x-\mu)^2}{\sigma^2}$, whereas the new values would be:

$$\begin{aligned} \sigma_{XY} &= \sqrt{\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}} \\ \mu_{XY} &= \frac{\mu_1\sigma_2^2+\mu_2\sigma_1^2}{\sigma_1^2+\sigma_2^2} \end{aligned} \tag{4}$$

iv)

Now it is possible to reformulate the whole function, as well as its proportionality, with this new terms:

$$f(x,y) = \frac{1}{2\pi\sigma_1^2\sigma_2^2}\exp\left(-\frac{1}{2}\frac{(x-\mu_{XY})^2}{\sigma_{XY}^2}\right) \propto \exp\left(-\frac{1}{2}\frac{(x-\mu_{XY})^2}{\sigma_{XY}^2}\right) \tag{5}$$

v)

Taking the simplified, proportional expression, it only remains to show that μ_{XY} is between μ_1 and μ_2, and that $\sigma_{XY}^2 < min(\sigma_1^2, \sigma_2^2)$. This can be done assuming $\mu_1 \leq \mu_2$

(the opposite would work the same way), and $\sigma_1^2, \sigma_2^2 \in \mathbb{R}^+$:

$$\begin{aligned}
\mu_{XY} &= \frac{\mu_1\sigma_2^2 + \mu_2\sigma_1^2}{\sigma_1^2+\sigma_2^2} \geq \frac{\mu_1\sigma_2^2 + \mu_1\sigma_1^2}{\sigma_1^2+\sigma_2^2} = \frac{\mu_1(\sigma_1^2+\sigma_2^2)}{\sigma_1^2+\sigma_2^2} = \mu_1 \iff \mu_{XY} \geq \mu_1 \\
\mu_{XY} &= \frac{\mu_1\sigma_2^2 + \mu_2\sigma_1^2}{\sigma_1^2+\sigma_2^2} \leq \frac{\mu_2\sigma_2^2 + \mu_1\sigma_1^2}{\sigma_1^2+\sigma_2^2} = \frac{\mu_2(\sigma_1^2+\sigma_2^2)}{\sigma_1^2+\sigma_2^2} = \mu_2 \iff \mu_{XY} \leq \mu_2 \\
\sigma_{XY} &= \sqrt{\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}} < \sqrt{\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2}} = \sqrt{\sigma_2^2} = \sigma_2 \iff \sigma_{XY} < \sigma_2 \\
\sigma_{XY} &= \sqrt{\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}} < \sqrt{\frac{\sigma_1^2\sigma_2^2}{\sigma_2^2}} = \sqrt{\sigma_1^2} = \sigma_1 \iff \sigma_{XY} < \sigma_1
\end{aligned} \tag{6}$$

□

Exercise 2

Let $p(x|\mu)$ be a univariate Gaussian $\mathcal{N}(\mu, \sigma^2)$ with unknown parameter mean, which is also assumed to follow a Gaussian $\mathcal{N}(\mu_0, \sigma_0^2)$. From the theory exposed before we have

$$p(\mu|X) = \frac{p(X|\mu)p(\mu)}{p(X)} = \frac{1}{\alpha}\prod_{k=1}^{N}\{p(x_k|\mu)p(\mu)\}$$

Where for a given training data set X, p(X) is a constant denoted as α. Write down the explicit expression for $p(\mu|X)$

i)

Since p, μ are assumed to be normally distributed, their PDFs can also be explicitly formulated:

$$\begin{aligned}
&\mu \sim \mathcal{N}(\mu_0, \sigma_0^2) \iff p(\mu) = \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left(-\frac{1}{2}\frac{(\mu-\mu_0)^2}{\sigma_0^2}\right) \\
&p(x_k|\mu) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\frac{(x_k-\mu)^2}{\sigma^2}\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\frac{(\mu-x_k)^2}{\sigma^2}\right) \\
&p(x_k|\mu)p(\mu) = \frac{1}{2\pi\sigma_0\sigma} \exp\left(-\frac{1}{2}\left(\frac{(\mu-x_k)^2}{\sigma^2} + \frac{(\mu-\mu_0)^2}{\sigma_0^2}\right)\right)
\end{aligned} \tag{7}$$

ii)

Which, as we already saw, corresponds to a function proportional to a normal PDF:

$$\begin{aligned} \sigma_H &= \sqrt{\frac{\sigma^2\sigma_0^2}{\sigma^2+\sigma_0^2}} \\ \mu_H &= \frac{\mu\sigma_0^2+\mu_0\sigma^2}{\sigma^2+\sigma_0^2} \\ p(x_k|\mu)p(\mu) &= \frac{1}{2\pi\sigma_0\sigma}\exp\left(-\frac{1}{2}\frac{(\mu-\mu_H)^2}{\sigma_H^2}\right) \end{aligned} \tag{8}$$

iii)

The important note here is that, since both x and μ are normally distributed, $p(x_k|\mu) \propto p(\mu|x_k)$, because the PDF is based on the euclidean (or L_2, or *radial*) distance between both of them, and this distance is symmetric (as it has to be in every norm). Therefore, and after re-normalizing it, **the expression can be directly reformulated as depending on $\boldsymbol{\mu}$**, and the conditioned PDF remains as follows:

$$\begin{aligned} p(\mu|X) = \frac{p(X|\mu)p(\mu)}{p(X)} &= \frac{1}{\alpha}\prod_{k=1}^{N}\{p(x_k|\mu)p(\mu)\} = \frac{1}{\alpha}\prod_{k=1}^{N}\{\frac{1}{2\pi\sigma_0\sigma}\exp\left(-\frac{1}{2}\frac{(\mu-\mu_H)^2}{\sigma_H^2}\right)\} \\ &= \frac{1}{\alpha(2\pi\sigma_0\sigma)^N}\exp\left(-\frac{1}{2\sigma_H^2}\sum_{k=1}^{N}\{(\mu-\mu_H)^2\}\right) \end{aligned} \tag{9}$$

Exercise 3

Show that, given a number of samples, N, the posterior $p(\mu|X)$ turns out to be also Gaussian, that is

$$p(\mu|X) = \frac{1}{\sigma_N\sqrt{2\pi}}\exp\left(-\frac{1}{2}\frac{(\mu-\mu_N)^2}{\sigma_N^2}\right)$$

with mean value

$$\mu_N = \frac{N\sigma_0^2\overline{x}_N+\sigma^2\mu_0}{N\sigma_0^2+\sigma^2}$$

and variance

$$\sigma_N^2 = \frac{\sigma^2\sigma_0^2}{N\sigma_0^2 + \sigma^2}$$

Where $\overline{x}_N = \frac{1}{N}\sum_{k=1}^{N} x_k$. *In the limit of large N, what happens to the mean value* μ_N *and to the standard deviation* σ_N*?*

i)

Assuming the samples aren't correlated to each other, and following the central limit theorem, the following is to be assumed:

$$\begin{aligned}
&p(\mu) = \mathcal{N}(\mu|\mu_0, \sigma_0^2) = \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left(-\frac{1}{2}\frac{(x-\mu_0)^2}{\sigma_0^2}\right) \\
&p(X|\mu,\sigma) = \prod_{k=1}^{N}\{p(x_k|\mu,\sigma)\} = \prod_{k=1}^{N}\{\mathcal{N}(x_k|\mu,\sigma)\} = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left(-\frac{1}{2\sigma^2}\sum_{k=1}^{N}\{(x_k-\mu)^2\}\right) \\
&p(X) = \frac{1}{N}
\end{aligned} \tag{10}$$

ii)

With this setup, and assuming a known σ, Bayes' theorem applies:

$$\begin{aligned}
p(\mu|X) &= \frac{p(X|\mu)p(\mu)}{p(X)} = \\
&= N\frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left(-\frac{1}{2\sigma^2}\sum_{k=1}^{N}\{(x_k-\mu)^2\}\right) \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left(-\frac{1}{2}\frac{(x-\mu_0)^2}{\sigma_0^2}\right) \\
&= N\frac{1}{(2\pi)^{\frac{N}{2}}\sigma^N\sigma_0} \exp\left(-\frac{1}{2\sigma^2}\sum_{k=1}^{N}\{(x_k-\mu)^2\} - \frac{1}{2}\frac{(x-\mu_0)^2}{\sigma_0^2}\right) \\
&= N\frac{1}{(2\pi)^{\frac{N}{2}}\sigma^N\sigma_0} \exp\left(-\frac{1}{2}\left(\frac{\sum_{k=1}^{N}\{(x_k-\mu)^2\}}{\sigma^2} + \frac{(\mu-\mu_0)^2}{\sigma_0^2}\right)\right)
\end{aligned}$$

Again, expanding the second term of the exponential function yields:

$$\frac{\sum_{k=1}^{N}\{(x_k-\mu)^2\}}{\sigma^2}+\frac{(\mu-\mu_0)^2}{\sigma_0^2}=\frac{\sigma_0^2\sum_{k=1}^{N}\{(x_k-\mu)^2\}+\sigma^2(\mu-\mu_0)^2}{\sigma^2\sigma_0^2}=$$

$$\frac{\sigma_0^2\sum_{k=1}^{N}\{(x_k^2-2x_k\mu+\mu^2)\}+\sigma^2(\mu^2-2\mu\mu_0+\mu_0^2)}{\sigma^2\sigma_0^2}=(\text{because }\sum_{k=1}^{N}\{x\}=N\overline{x})$$

$$=\frac{\mu^2(N\sigma_0^2+\sigma^2)+\sigma_0^2(\sum_{k=1}^{N}\{x_k^2\}-2\mu N\overline{x})+\sigma^2\mu_0^2-2\sigma^2\mu\mu_0}{\sigma^2\sigma_0^2}$$

$$=\frac{\mu^2(N\sigma_0^2+\sigma^2)-2\mu(\sigma^2\mu_0^2+\sigma_0^2N\overline{x})+\sigma_0^2\sum_{k=1}^{N}\{x_k^2\}+\sigma^2\mu_0}{\sigma^2\sigma_0^2}$$

$$=\frac{\mu^2-2\mu(\frac{\sigma^2\mu_0^2+\sigma_0^2N\overline{x}}{N\sigma_0^2+\sigma^2})+\sigma_0^2\sum_{k=1}^{N}\{x_k^2\}+\sigma^2\mu_0}{\frac{\sigma^2\sigma_0^2}{N\sigma_0^2+\sigma^2}}$$

$$=\frac{(\mu-\frac{\sigma^2\mu_0+\sigma_0^2N\overline{x}}{N\sigma_0^2+\sigma^2})^2}{\frac{\sigma^2\sigma_0^2}{N\sigma_0^2+\sigma^2}}=\frac{(\mu-\mu_N)^2}{\sigma_N^2}$$

iii)

In the limit of large N, the initial assumptions μ_0 and σ_0 become irrelevant, and the final distribution becomes a certain scalar(see the lecture slides, page 11 for more information):

$$\lim_{N\to\infty}\mu_N=\overline{x}_N$$
$$\lim_{N\to\infty}\sigma_N^2=0$$

vspace5mm

Exercise 4

Plot the posterior distribution $p(\mu|X)$ from Exercise 3 in one graph for various N. The largest N should be at least as large as $N=100$. To compute $\overline{x}_N$ generate data $X=\{x_1,...,x_N\}$ using a pseudorandom number generator following a Gaussian pdf with mean value $\mu=2$ and variance $\sigma^2=4$. The mean value is assumed to be unknown and the prior pdf is also a Gaussian with $\mu_0=0$ and $\sigma_0^2=8$. Also include the prior in this plot and describe what happens when increasing N.

i)

See the fernandez_blatt4.py Python2 file for the details:

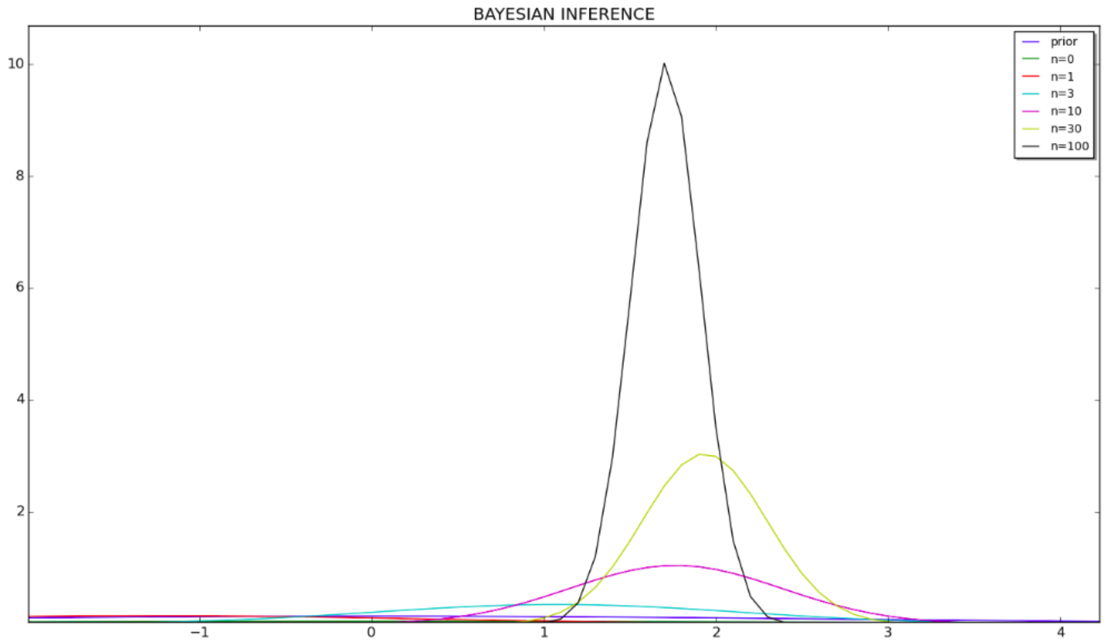

Figure 1: inference of $\mathcal{N}(x|2,4)$, starting by $\mathcal{N}(x|0,8)$