Pattern and Speech Recognition WS2015-16
Exercise 4

Atanas Poibrenski(2554135), Marimuthu Kalimuthu(2557695), Furkat Kochkarov(2557017)

November 27, 2015

Probability Theory

1. A,B $->$ two events

   P(A).P(B) = $P(A \cap B)$; $P(A|B)$ = P(A)
   Yes they are equivalent, since independence is assumed.

   $$\begin{aligned} P(A \cap B) &= P(A, B) \\ &= P(A).P(B|A) \end{aligned}$$

   If A & B are independent,

   $$= P(A).P(B)$$

   Also

   $$\boxed{P(A|B) = P(A)}$$

2. Bayes Law

   $$\begin{aligned} P(A \cap B) &= P(A).P(B|A) \\ &= P(B).P(A|B) \\ P(A|B) &= \frac{P(B|A)P(A)}{P(B)} \end{aligned}$$

   which is Bayes law.

3. **Bonus:**

$$\mathbb{E}\left[X+Y\right] = \mathbb{E}\left[X\right] + \mathbb{E}\left[Y\right]$$

$$\begin{aligned}
\mathbb{E}\left[X\right] &= \int_{-\infty}^{\infty} x f(x) dx \\
&= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx \\
&= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{0} x e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx + \frac{1}{\sigma\sqrt{2\pi}} \int_{0}^{\infty} x e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx \\
&= \mu
\end{aligned}$$

similarly for

$$\mathbb{E}\left[Y\right] = \mu$$

$$\boxed{\mathbb{E}\left[X+Y\right] = 2\mu}$$

$$Var[X+Y] = Var[X] + Var[Y]$$

$$\begin{aligned}
Var[X] &= \mathbb{E}\left[(x-\mu)^2\right] \\
\mathbb{E}\left[(x-\mu)^2\right] &= \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx \\
&= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (x-\mu)^2 e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx
\end{aligned}$$

on evaluating the integral, we get

$$\begin{aligned}
&= 0 + \sigma^2 .1 \\
&= \sigma^2
\end{aligned}$$

similarly,

$$Var[Y] = \sigma^2$$

Hence,

$$\boxed{Var[X+Y] = 2\sigma^2}$$

MLE for Poisson Distribution

1. Likelihood

$$\begin{aligned}
L(\theta; k) &= \prod_{i=1}^{n} \frac{e^{-\theta}\theta^{k_i}}{k_i!} \\
&= e^{-n\theta}\theta
\end{aligned}$$

See ‘poisson_likelihood.m’

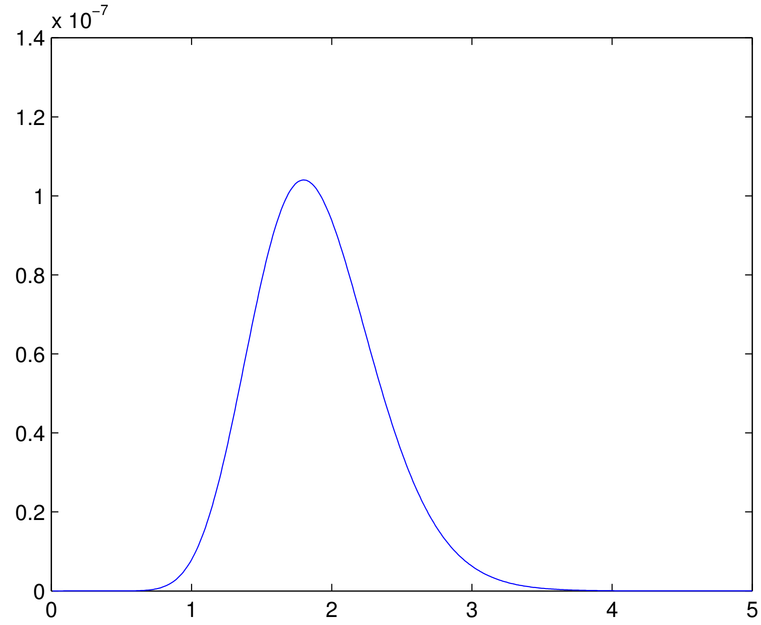

Figure 1: Poisson likelihood (for $\theta \in (0, 5]$)

2. See ‘poisson_loglikelihood.m’

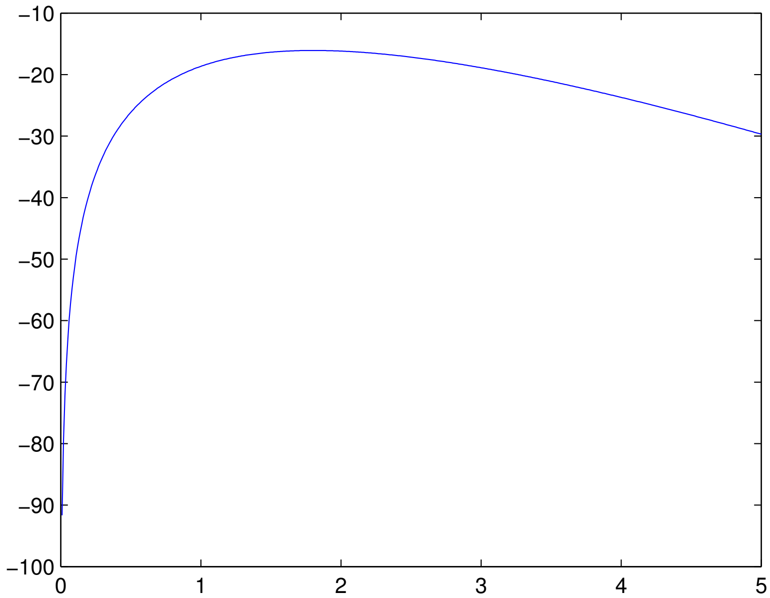

Figure 2: Poisson log-likelihood (for $\theta \in (0, 5]$)

Here we notice that since log is a monotonically increasing function, it achieves maximum at the same θ as in the likelihood plot. So, we can use either of them to find the maximum.

3. Here we use log-likelihood

$$L(\theta|x) = \sum_{i=1}^{n}(x_i ln\theta) - n\theta$$
$$\frac{\partial L(\theta|x)}{\partial\theta} = 0$$
$$\frac{\sum_{i=1}^{n} x_i}{\theta} - n = 0$$
$$\theta_{max} = \frac{\sum_{i=1}^{n} x_i}{n}$$

In our case, x=0,2,2,1,2,1,3,1,1,5

$$\boxed{\theta_{max} = 1.8}$$

MLE for Normal Distribution

1. After importing Iris-setosa, Iris-versicolor, we get the following 1-D plot (histogram)

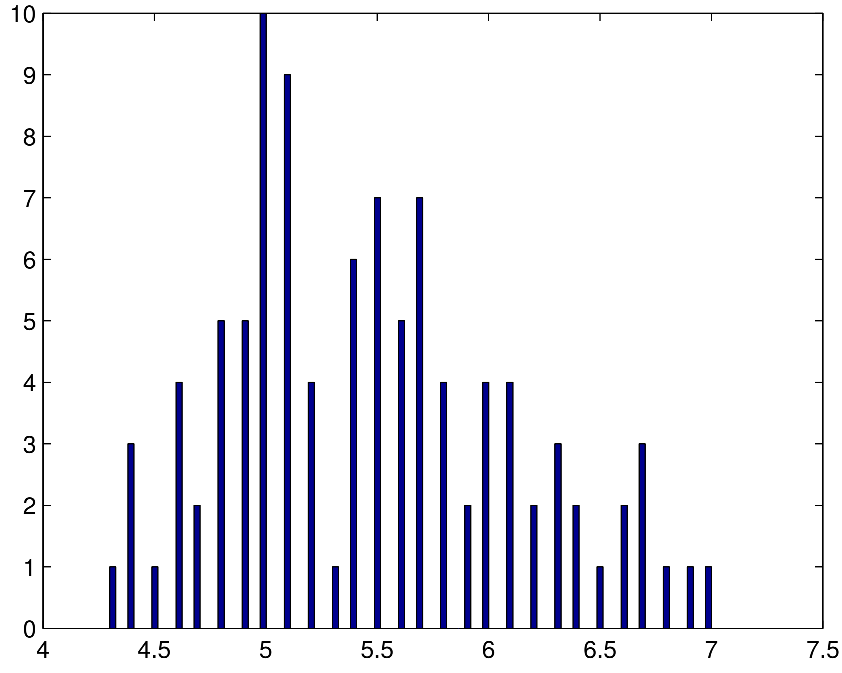

Figure 3: Iris-setosa and Iris-versicolor

2. Maximize the likelihood

$$p_\theta(x_i) = \frac{1}{b-a} \qquad if(a \leq x_i \leq b)$$

To maximize $p_\theta(x_i)$ we need to minimize (b-a). The smallest interval which contains all the points. That is a=min(x) and b=max(x)

.

In the case of the Iris dataset, a=4.3 and b=7

3. p_θ is a probability density.
   $p_\theta(x_i) = \frac{1}{2}.f_{\mu_1,\sigma 1}(x_i) + \frac{1}{2}.f_{\mu_2,\sigma 2}(x_i)$

   The normal distribution is a probability density because $\int p(x)dx = 1$
   p_θ is a probability density because,

   $=> \frac{1}{2}.1 + \frac{1}{2}.1 = 1$

4. See 'normal_likelihood.m' and 'normal_loglikelihood.m'
   We prefer log-likelihood since it is numerically stable. Since the likelihood gives very small probabilities (Figure 1), it is easier to compute and represent log-likelihood where the product becomes summation.

5. We use log-likelihood to estimate the parameter θ_{max} for the reason stated above. We used 'fminsearch' in matlab with the initial guess(6.0, 0.6, 5.0, 0.5) in order to find
   $\theta_{max} = [5.47, 0.63, 5.25, 0.5]$

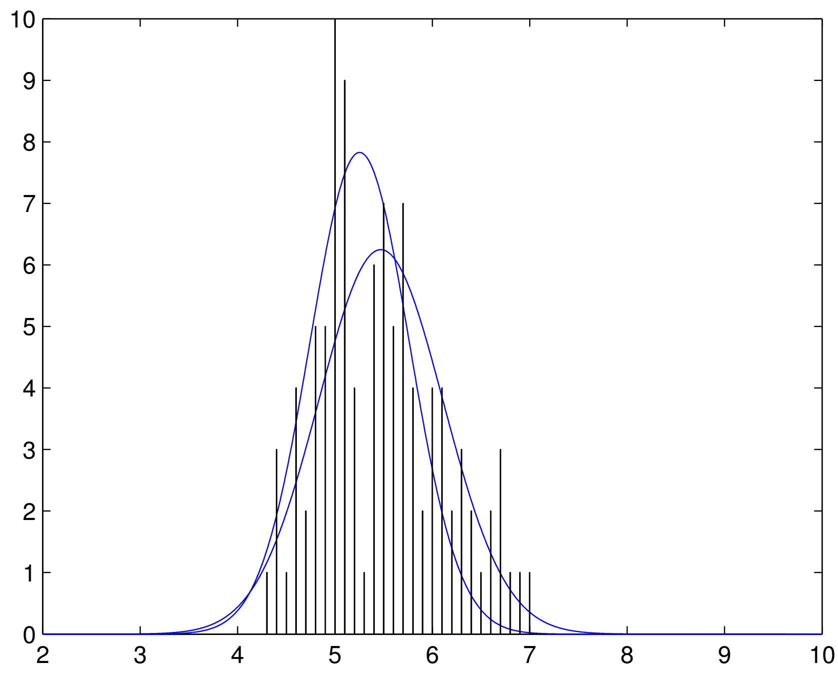

Figure 4: Normal distribution

See 'drawgaussian.m'