SOLUTIONS FOR PATTERN CLASSIFICATION CH.3

Building C303, Harbin Institute of Technology, Shenzhen University Town, Xili, Shenzhen,Guangdong Province, P.R. China
Yonghui Wu, Yaoyun Zhang; yonghui.wu@gmail.com

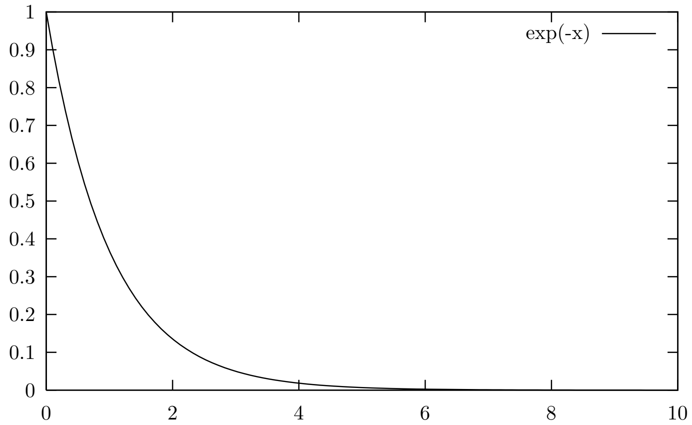
Problem 1

- a

When $\theta = 1$

$$p(x|\theta) = \begin{cases} e^{-x}, & x \geq 0 \\ 0 & others \end{cases}$$

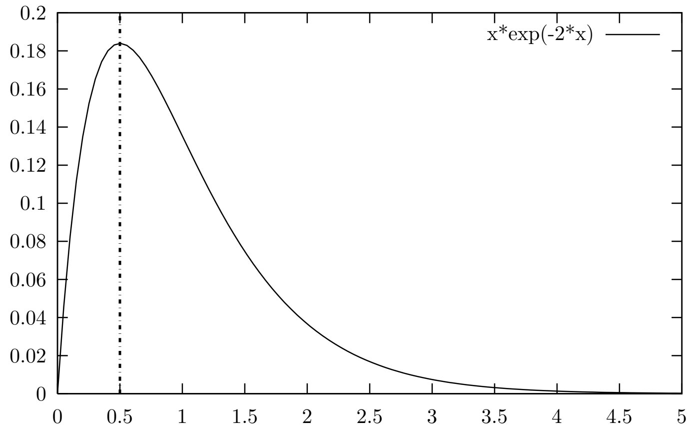
Plot as follows:



When $x = 2$ we get:

$$p(x|\theta) = \begin{cases} \theta e^{-2\theta}, & x \geq 0 \\ 0 & others \end{cases}$$

Plot as follows:



- b

$x_1, \cdots, x_n$ $p(x|\theta)$, and x_i are independently. Then we define the log-likelihood as follows:

$$l(\theta) = \sum_{i=1}^{n} \ln p(x_i|\theta) = \sum_{i=1}^{n} (\ln\theta - \theta x_i)$$

Then we get:

$$\frac{dl(\theta)}{d\theta} = \sum_{i=1}^{n} (\frac{1}{\theta} - x_i) = \frac{n}{\theta} - \sum_{i=1}^{n} x_i = 0$$

Thus:

$$\hat{\theta} = \frac{1}{n\sum_{i=1}^{n} x_i}$$

- c

  $\theta = 1$, when n is large enough we get:

$$\begin{aligned}\lim_{n\to\infty} (\frac{1}{n}\sum_{i=1}^{n} x_i) &= \int_{-\infty}^{\infty} xe^{-x}dx \\ &= -(x+1)e^{-x}|_0^{\infty} = 1\end{aligned}$$

Problem 2

- a x have a uniform density, when $\theta \geq \max(\mathcal{D})$ define the likelihood as:

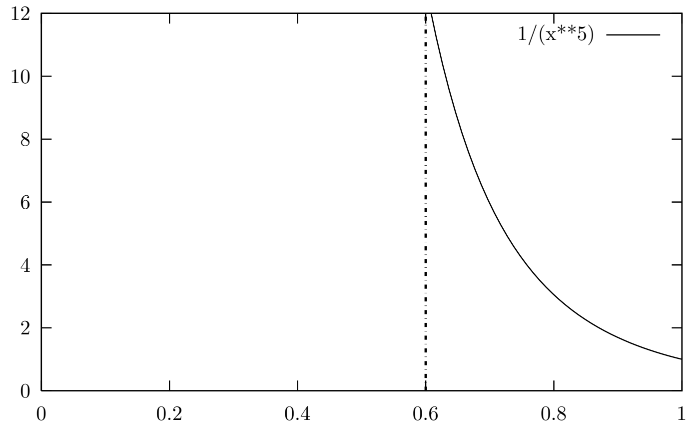
$$\begin{aligned}l(\mathcal{D}|\theta) &= \prod_{i=1}^{n} p(x_i|\theta) \\ &= \prod_{i=1}^{n} \frac{1}{\theta} = \frac{1}{\theta^n}\end{aligned}$$

  $l(\theta)$ is monotonically decreasing, the $\hat{\theta}$ that maximizes the $l(\theta)$ minimize the θ. As $\theta \geq \max(\mathcal{D})$, the $l(\theta)$ is maximized in $\hat{\theta} = \max(\mathcal{D})$

- b

  When $n = 5$, From a we know that the likelihood of $p(\mathcal{D}|\theta)$ is: $\frac{1}{\theta^5}$, the $\hat{\theta} = \max(\mathcal{D}) = \max(x_i)$. Thus we can get $\hat{\theta}$, and need not to know the values of the other four points as soon as we know the $\max x_i = 0.6$. When $\theta < 0.6,\ \max(x_i) > \theta,\ so\ p(\mathcal{D}|\theta) = 0$.

Plot:



Problem 4

- a

Because X is a d-dimensional binary vector with a multivariate Bernoulli distribution:

$$P(X|\theta) = \prod_{i=1}^{d} \theta_i^{x_i}(1-\theta_i)^(1-x_i)$$

hence,

$$P(X^1, ..., X^n|\theta) = \prod_{k=1}^{n}\prod_{i=1}^{d} \theta_i^{x_i^k}(1-\theta_i)(1-x_i^k)$$

And the likelihood function:

$$l(\theta) = \sum_{k=1}^{n}\sum_{i=1}^{d}\left[x_i^k \ln\theta_i + (1-x_i^k)\ln(1-x_i)\right]$$

$$\frac{\partial l(\theta)}{\partial\theta_i} = \sum_{k=1}^{n}\left[\frac{x_i^k}{\theta_i} - \frac{1-x_i^k}{1-\theta_i}\right] = \sum_{k=1}^{n}\frac{x_i^k-\theta_i}{\theta_i(1-\theta_i)}$$

thus, we get

$$\sum_{k=1}^{n}(x_i^k - \theta_i) = 0$$

$$\theta_i = \frac{1}{n}\sum_{k=1}^{n} x_i^k$$

Then we can get its vector form as follows:

$$\theta_i = \frac{1}{n}\sum_{k=1}^{n} X_k$$

Problem 16

- a

  Since A and B are nonsingular matrices of the same order,

$$\begin{aligned}
(A^{-1}+B^{-1})[A(A+B)^{-1}B] &= (1+B^{-1}A)(A+B)^{-1}B \\
&= (A+B)^{-1}B + B^{-1}A(A+B)^{-1}B \\
&= B^{-1}B(A+B)^{-1}B + B^{-1}A(A+B)^{-1}B \\
&= B^{-1}(B+A)(A+B)^{-1}B \\
&= B^{-1}B \\
&= I
\end{aligned}$$

  Left multiply $(A^{-1}+B^{-1})^{-1}$ on both sides of the above equation, we get:

$$(A^{-1}+B^{-1})^{-1} = A(A+B)^{-1}B$$

  Analogously,

$$\begin{aligned}
B(A+B)^{-1}A(A^{-1}+B^{-1}) &= B(A+B)^{-1}(I+AB^{-1}) \\
&= B(A+B)^{-1} + B(A+B)^{-1}AB^{-1} \\
&= B(A+B)^{-1}BB^{-1} + B(A+B)^{-1}AB^{-1} \\
&= B(A+B)^{-1}(B+A)B^{-1} \\
&= BB^{-1} \\
&= I
\end{aligned}$$

  Right multiply $(A^{-1}+B^{-1})^{-1}$ on both sides of the above equation, we get:

$$(A^{-1}+B^{-1})^{-1} = B(A+B)^{-1}A$$

- b

  If these matrices are not square, then they are both nonsingular, A^{-1} and B^{-1} won't exist, which means the above equation won't hold, so, A and B must be square.

- c

  From (41), $\sum_n^{-1} = n\sum^{-1} + \sum_0^{-1}$, we can get

$$\Sigma_n = (n\Sigma^{-1}+\Sigma_0^{-1})^{-1} = (\Sigma_0^{-1}+n\Sigma^{-1})^{-1}$$

  Making use of the matrix identity in (a), we have

$$\Sigma_n = \Sigma_0(\Sigma_0+\frac{1}{n}\Sigma)^{-1}(\frac{1}{n}\Sigma)$$

  Thus, we get Eqs.(45) similarly, we have

$$\Sigma_n = (n\Sigma^{-1}+\Sigma_0^{-1})^{-1} = (\frac{1}{n}\Sigma)(\Sigma_0+\frac{1}{n}\Sigma)^{-1}\Sigma_0$$

From (42), $\Sigma_n^{-1}\mu_n = n\Sigma^{-1}\widehat{\mu}_n + \Sigma_0^{-1}\mu_0$, Left multiply the above equation by Σ_n:

$$\begin{aligned}
\mu_n &= \Sigma_n(n\Sigma^{-1}\widehat{\mu}_n + \Sigma_0^{-1}\mu_0) \\
&= \Sigma_n n\Sigma^{-1}\widehat{\mu}_n + \Sigma_n\Sigma_0^{-1}\mu_0 \\
&= \Sigma_0(\Sigma_0 + \frac{1}{n}\Sigma)^{-1}(\frac{1}{n}\Sigma)n\Sigma^{-1}\widehat{\mu}_n + \frac{1}{n}\Sigma(\frac{1}{n}\Sigma + \Sigma_0)^{-1}\Sigma_0\Sigma_0^{-1}\mu_0 \\
&= \Sigma_0(\Sigma_0 + \frac{1}{n}\Sigma)^{-1}(\frac{1}{n}\Sigma)\widehat{\mu}_n + \frac{1}{n}\Sigma(\frac{1}{n}\Sigma + \Sigma_0)^{-1}\mu_0
\end{aligned}$$

Which is Eqs.46.