EE E6887 (Statistical Pattern Recognition)
Solutions for homework 3

P.1 Assume the training samples are distributed in the unit hypercube in the d-dimensional feature space, R^d. Compute $l_d(p)$, the length of a hypercube edge in d dimensions that contains the fraction p of points ($0 \leq p \leq 1$). To better appreciate the implications of your result, compute $l_5(0.01)$, $l_5(0.1)$, $l_{20}(0.01)$ and $l_{20}(0.1)$.

Answer: If the training samples are uniformly distributed in the d-dimensional hypercube, then we have:

$$p = (l_d(p))^d$$
$$\Longrightarrow l_d(p) = \sqrt[d]{p}$$

Then

$$\begin{aligned} l_5(0.01) &= 0.3981 \\ l_5(0.1) &= 0.6310 \\ l_{20}(0.01) &= 0.7943 \\ l_{20}(0.1) &= 0.8913 \end{aligned}$$

P.2 Using the results given in Table 3.1, show that the maximum likelihood estimate for parameter θ of a Rayleigh distribution is given by:

$$\hat{\theta} = \frac{1}{\frac{1}{n}\sum_{k=1}^{n} x_k^2}$$

Answer:

$$\begin{aligned} l &= \sum_{k=1}^{n} \log P(x_k|\theta) \\ &= \sum_{k=1}^{n} \log \left(2\theta x_k e^{-\theta x_k^2}\right) \\ &= n \log 2\theta + \sum_{k=1}^{n} \log x_k - \theta \sum_{k=1}^{n} x_k^2 \end{aligned}$$

thus

$$\frac{dl}{d\theta} = \frac{n}{\theta} - \sum_{k=1}^{n} x_k^2 = 0$$

$$\Longrightarrow \hat{\theta} = \frac{1}{\frac{1}{n}\sum_{k=1}^{n} x_k^2}, \qquad (x_k \geq 0, k = 1, \ldots, n)$$

Further to see $\hat{\theta} = \frac{1}{\frac{1}{n}\sum_{k=1}^{n} x_k^2}$ is the sufficient statistics, we write $P(\mathcal{D}|\theta)$ as follows:

$$\begin{aligned} P(\mathcal{D}|\theta) &= \prod_{k=1}^{n} \left(2\theta x_k e^{-\theta x_k^2}\right) \\ &= \prod_{k=1}^{n} 2x_k \prod_{k=1}^{n} \theta e^{-\theta x_k^2} \\ &= \left(\prod_{k=1}^{n} 2x_k\right)\left(\theta^n e^{-\theta \sum_{k=1}^{n} x_k^2}\right) \\ &= P(\mathcal{D}) g(s, \theta) \end{aligned}$$

So $\sum_{k=1}^{n} x_k^2$, the denominator of $\hat{\theta}$, is one form of the sufficient statistics.

P.3 Option A: Consider data $\mathcal{D} = \{(2,3)^t, (3,1)^t, (5,4)^t, (4,*)^t, (*,6)^t\}$ sampled from a two-dimensional uniform distribution:

$$p(\mathbf{x}) \sim U(\mathbf{x}_l, \mathbf{x}_u) = \begin{cases} \frac{1}{|x_{u1}-x_{l1}||x_{u2}-x_{l2}|}, & x_{l1} \leq x_1 \leq x_{u1} \;\&\; x_{l2} \leq x_2 \leq x_{u2} \\ \epsilon, & otherwise \end{cases}$$

where * represents missing feature values and ϵ is a very small positive constant that can be neglect when normalizing the density within the above bounds.

(a) Start with an initial estimate $\theta^0 = (\mathbf{x}_l^t, \mathbf{x}_u^t)^t = (0, 0, 10, 10)^t$ and analytically calculate $Q(\theta; \theta^0)$ – the E-step in the EM algorithm.

(b) Find the θ that maximize your $Q(\theta; \theta^0)$ – the M-step. You may make some simplifying assumptions.

(c) Plot your data and the bounding rectangle

(d) Without having to iterate further, state the estimate of θ that would result after convergence of the EM algorithm.

Answer:

$$p(x) \sim U(x_l, x_u) = I(x_l \leq x \leq x_u)A(x_l, x_u)$$

where

$$A(x_l, x_u) = \frac{1}{|x_{u1} - x_{l1}||x_{u2} - x_{l2}|}$$

and

$$I(x) = 1 \textbf{ if } x = \textbf{true}, I(x) = 0, \textbf{otherwise}$$

In E-step, we compute the $Q(\theta, \theta^0)$, where $\theta = (x_{l1}, x_{l2}, x_{u1}, x_{u2})$, and $\theta^0 = (0, 0, 10, 10)$

Let, the hidden variables, $h = (x_{42}, x_{51})$ and $X = \{x_i, i = 1, 2, \ldots, 5\}\backslash h$.

$$Q(\theta, \theta^0) = \int_h ln(p(X, h|\theta))p(h|X, \theta^0)dh$$

$$= \sum_{k=1}^{3} ln(p(x_i|\theta)) + \int ln(p(x_4|\theta))p(x_{42}|X, \theta^0)dx_{42} + \int ln(p(x_5|\theta))p(x_{51}|X, \theta^0)dx_{51}$$

$$= \sum_{k=1}^{3} ln(p(x_i|\theta)) + C_4 \int ln(p(x_4|\theta))p(x_4|x_{41} = 4, \theta^0)dx_{42}$$

$$+C_5 \int ln(p(x_5|\theta))p(x_5|x_{52} = 6, \theta^0)dx_{51}$$

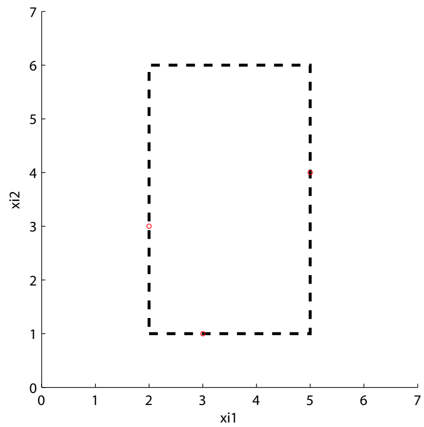
$$= 3ln(A(\theta)) + 10C_4 ln(A(\theta))A(\theta^0) + 10C_5 ln(A(\theta))A(\theta^0)$$

$$= ln(A(\theta))(3 + 10C_4 A(\theta^0) + 10C_5 A(\theta^0))$$

$$max_\theta Q(\theta, \theta^0) = max_\theta A(\theta) = min_\theta |x_{u1} - x_{l1}||x_{u2} - x_{l2}|$$

The minimum θ can be found by looking for the smallest bounding box that contains all the points, therefore $\theta = (2, 1, 5, 6)$

The subsequent iterations will give us the same expression for $Q(\theta, \theta^0)$, therefore the maximized θ will remain the same.



P.3 Option B: Let $p(x) \sim U(0,a)$ be uniform from 0 to a, and let a Parzen window be defined as $\varphi(x) = e^{-x}$ for $x > 0$ and 0 for $x \leq 0$.

(a) Show that the mean of such a Parzen-window estimate is given by:

$$\bar{p}_n(x) = \begin{cases} 0, & x < 0 \\ \frac{1}{a}(1 - e^{-x/h_n}), & 0 \leq x \leq a \\ \frac{1}{a}(e^{a/h_n} - 1)e^{-x/h_n}, & a \leq x \end{cases}$$

(b) Plot $\bar{p}_n(x)$ versus x for $a = 1$ and $h_n = 1, 1/4$, and $1/16$.

Answer:

(a) Since $p(x) \sim U(0,a)$. Thus the training samples $x_1, \ldots, x_n \in [0,a]$ and, for all x_i, the volume, V_n is given by:

$$V_n = \int e^{(x-x_i)/h_n} = h_n$$

Case 1: when $x \geq a$, $x - x_i \geq 0$, for $i = 1, \ldots, n$. We have:

$$\begin{aligned} p_n(x) &= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_n}\varphi(\frac{x-x_i}{h_n}) \\ &= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_n}e^{-(\frac{x-x_i}{h_n})} \end{aligned}$$

So in this case

$$
\begin{aligned}
\overline{p}_n(x) &= E\{p_n(x)\} \\
&= \frac{1}{n}\sum_{i=1}^{n} E\{\frac{1}{h_n}e^{-(\frac{x-x_i}{h_n})}\} \\
&= \int_0^a \frac{1}{ah_n}e^{-(\frac{x-v}{h_n})}dv \\
&= \frac{1}{a}e^{-x/h_n}(e^{a/h_n}-1)
\end{aligned}
$$

Case 2: when $x < 0$, $x - x_i < 0$, for $i = 1, \ldots, n$, and $p_n(x) = 0$. So $\overline{p}_n(x) = 0$.

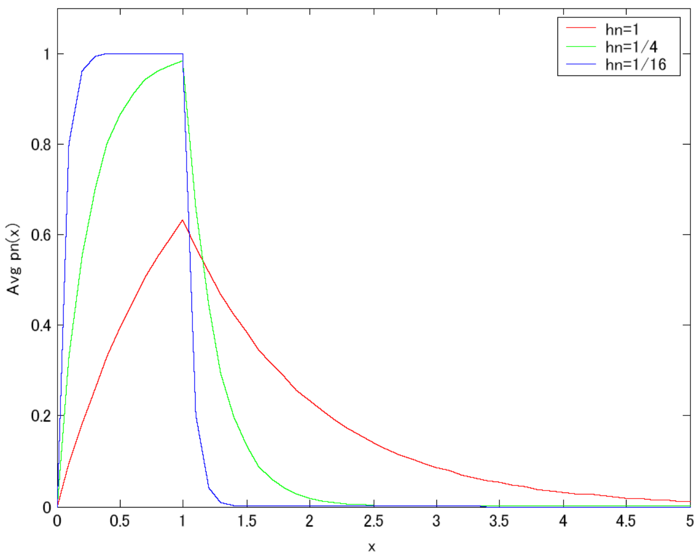
Case 3: when $0 \le x \le a$, we have:

$$
\begin{aligned}
\overline{p}_n(x) &= \int_0^x \frac{1}{ah_n}e^{-(\frac{x-v}{h_n})}dv \\
&= \frac{1}{a}e^{-x/h_n}(e^{x/h_n}-1) \\
&= \frac{1}{a}(1-e^{-x/h_n})
\end{aligned}
$$

So in summary:

$$
\overline{p}_n(x) = \begin{cases} 0, & x < 0 \\ \frac{1}{a}(1-e^{-x/h_n}), & 0 \le x \le a \\ \frac{1}{a}(e^{a/h_n}-1)e^{-x/h_n}, & a \le x \end{cases}
$$

(b) The following figure shows $\overline{p}_n(x)$ versus x for $a = 1$, and $h_n = 1, 1/4, 1/16$ respectively.



hn=1
hn=1/4
hn=1/16
Avg pn(x)
x