Tik-61.231 Principles of Pattern Recognition

Answers to exercise 3: 9.10.2000

1. The poisson distribution: $p(x|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}, x \geq 0$.

   The maximum likelihood estimate for the parameter λ:
   $p(H|\lambda) = \prod_{k=1}^{n} p(x_k|\lambda)$, where $H = \{x_1, \dots, x_n\}$ is the sample set.

   The maximum likelihood estimate can be found by maximizing $p(H|\lambda)$. The same solution can also be found by maximizing the functions natural logarithm, as the logarithm function is monotonically increasing. The maximum can be found by setting the derivative regarding λ to zero,

   $\frac{\delta}{\delta\lambda} \ln\{p(H|\lambda)\} = \frac{\delta}{\delta\lambda} \sum_{k=1}^{n} (x_k \ln(\lambda) - \lambda - \ln(x_k!)) = \sum_{k=1}^{n} (x_k \frac{1}{\lambda} - 1) = \frac{1}{\lambda} \sum_{k=1}^{n} x_k - n = 0$
   $\Rightarrow \hat{\lambda} = \frac{1}{n} \sum_{k=1}^{n} x_k$

   For an unbiased estimate $E\{\hat{\theta}\} = \theta$.

   $E\{\hat{\lambda}\} = E\{\frac{1}{n} \sum_{k=1}^{n} x_k\} = \frac{1}{n} \sum_{k=1}^{n} E\{x_k\} = E\{x_k\} = \lambda$, as for the Poisson distribution $E\{x\} = \lambda$.

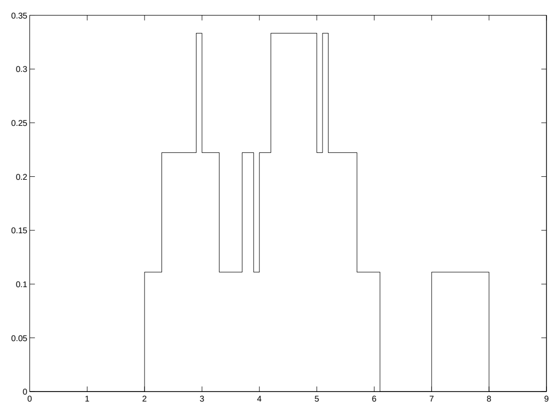
   Thus $\hat{\lambda}$ is unbiased.

2. Parzen window $\Theta(x) = \Theta\left(\frac{\overline{x} - \overline{x^i}}{h_n}\right) = \begin{cases} 1, \text{ if } \overline{x} \text{ in the hypercuve center } \overline{x^i}, \text{ side length } h_n \\ 0, \text{ otherwise} \end{cases}$

   $k_n = \sum_{i=1}^{n} \Theta\left(\frac{\overline{x} - \overline{x^i}}{h_n}\right)$ is the number of samples in the hypercube and n the total number of samples.

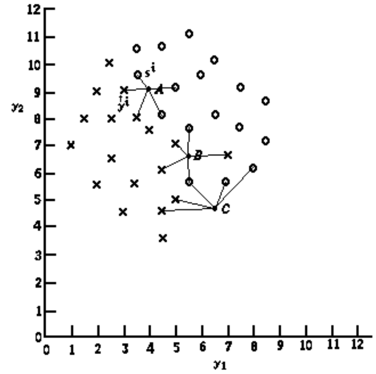
   $p_n(\overline{x}) = \frac{k_n/n}{V_n} = \frac{1}{nV_n} \sum_{i=1}^{n} \Theta\left(\frac{\overline{x} - \overline{x^i}}{h_n}\right)$, where V_n is the volume of the hypercube.

   Now $\overline{x^i} = y^{(i)}$, $h_n = 1$, $V_n = 1$ and $n = 9$. Thus



3. Point X is classified according to it's 5 nearest neighbours. The distance metric used is the Euclidian distance and the decision rule the voting result of the 5 nearet neighbours.

Point	Votes for O	Votes for X	Resulting class
A	3	2	O
B	2	3	X
C	3	2	O



4. **a)** $y^{(i)} = [y_1^{(i)}\ y_2^{(i)}]^T, s^{(i)} = [s_1^{(i)}\ s_2^{(i)}]^T$. A point $\hat{y} = [y_1\ y_2]$ is on the decision boundary, if

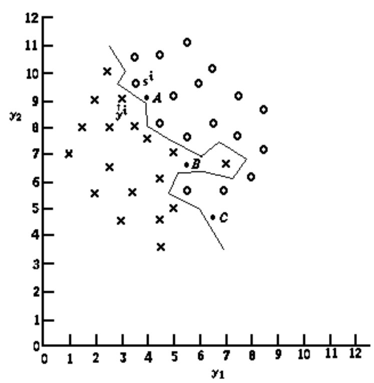
$$d(\hat{y}, y^{(i)}) = d(\hat{y}, s^{(j)})$$
$$\sqrt{(\hat{y}_1 - y_1^{(i)})^2 + (\hat{y}_2 - y_2^{(i)})^2} = \sqrt{(\hat{y}_1 - s_1^{(i)})^2 + (\hat{y}_2 - s_2^{(i)})^2}$$

$\Leftrightarrow \hat{y}_1 2(s_1^{(j)} - y_1^{(i)}) + \hat{y}_2 2(s_2^{(j)} - y_2^{(i)}) + (y_1^{(i)})^2 + (s_1^{(i)})^2 + (y_2^{(i)})^2 + (s_2^{(i)})^2 = 0$

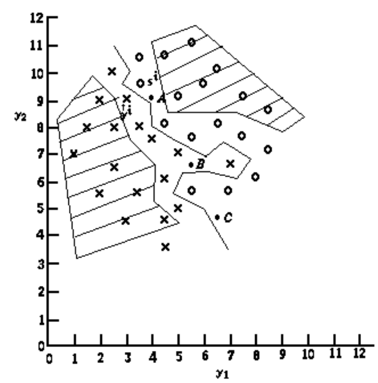
and by setting $a = 2(s_1^{(j)} - y_1^{(i)})$, $b = 2(s_2^{(j)} - y_2^{(i)})$ and $c = (y_1^{(i)})^2 + (s_1^{(i)})^2 + (y_2^{(i)})^2 + (s_2^{(i)})^2$ it is clear that the resulting equation $a\hat{y}_1 + b\hat{y}_2 + c = 0$ represents a line. By taking the points from the figure,

$$\begin{cases} s_1^{(j)} = 3.6 \\ s_2^{(j)} = 9.5 \\ y_1^{(j)} = 3.0 \\ y_2^{(j)} = 9.0 \end{cases} \Rightarrow \begin{cases} a = 1.2 \\ b = 1.0 \\ c = -13.12 \end{cases} \Rightarrow \hat{y}_2 = -1.2\hat{y}_1 + 13.12$$

b)



c)



5. A 1-NN Classifier is created by using a training set $H_n = \{\overline{x}_1, \ldots, \overline{x}_n\}$ and it is tested with the vector $\overline{x}$. $\overline{x}$ belongs to the class ω and its nearest neighbour $\overline{x}'$ belongs to class θ. The error probability for $\overline{x}$ is (Schalkoff page 79):

$$e_{1NNR}(\overline{x}, \overline{x}') = P(\omega \neq \theta | \overline{x}, \overline{x}') = \textstyle\sum_i P(\omega = \omega_i | \overline{x}) P(\theta \neq \omega_i | \overline{x}')$$

Now $\overline{x} = \overline{x}'$ as the training set is also used as a test set.

$\Rightarrow \forall i P(\theta \neq \omega_i | \overline{x}') = 0$, if $\omega = \omega_i \Rightarrow e_{1NNR} = 0$

This is obviously an useless estimate for the recognition performance.