EE3731C Tutorial - Pattern Recognition 2

Department of Electrical and Computer Engineering

1. $p_h(x) = \frac{1}{4}\sum_{n=1}^{4} \frac{1}{\sqrt{2\pi h^2}} e^{-\frac{(x_n-x)^2}{2h^2}}$.

   (a) Plugging $x = 4$, $h = 1$, and the data, we get

   $$\begin{aligned} p_1(x=4) &= \frac{1}{4\sqrt{2\pi}}\left[e^{-\frac{(2-4)^2}{2}} + e^{-\frac{(5-4)^2}{2}} + e^{-\frac{(7-4)^2}{2}} + e^{-\frac{(8-4)^2}{2}}\right] \\ &= \frac{1}{4\sqrt{2\pi}}\left[e^{-2} + e^{-\frac{1}{2}} + e^{-9/2} + e^{-8}\right] \\ &= 0.0751318 \end{aligned}$$

   Plugging $x = 4$, $h = 2$, and the data, we get

   $$\begin{aligned} p_2(x=4) &= \frac{1}{4\sqrt{8\pi}}\left[e^{-\frac{(2-4)^2}{8}} + e^{-\frac{(5-4)^2}{8}} + e^{-\frac{(7-4)^2}{8}} + e^{-\frac{(8-4)^2}{8}}\right] \\ &= 0.0971931 \end{aligned}$$

   Plugging $x = 4$, $h = 3$, and the data, we get

   $$\begin{aligned} p_3(x=4) &= \frac{1}{4\sqrt{18\pi}}\left[e^{-\frac{(2-4)^2}{18}} + e^{-\frac{(5-4)^2}{18}} + e^{-\frac{(7-4)^2}{18}} + e^{-\frac{(8-4)^2}{18}}\right] \\ &= 0.0919010 \end{aligned}$$

   (b) Plugging $x = 6$, $h = 3$, and the data, we get

   $$\begin{aligned} p_3(x=6) &= \frac{1}{4\sqrt{2\pi}}\left[e^{-\frac{(2-6)^2}{2}} + e^{-\frac{(5-6)^2}{2}} + e^{-\frac{(7-6)^2}{2}} + e^{-\frac{(8-6)^2}{2}}\right] \\ &= 0.134517 \end{aligned}$$

   Plugging $x = 6$, $h = 2$, and the data, we get

   $$\begin{aligned} p_3(x=6) &= \frac{1}{4\sqrt{8\pi}}\left[e^{-\frac{(2-6)^2}{8}} + e^{-\frac{(5-6)^2}{8}} + e^{-\frac{(7-6)^2}{8}} + e^{-\frac{(8-6)^2}{8}}\right] \\ &= 0.1250115 \end{aligned}$$

   Plugging $x = 6$, $h = 3$, and the data, we get

   $$\begin{aligned} p_3(x=6) &= \frac{1}{4\sqrt{18\pi}}\left[e^{-\frac{(2-6)^2}{18}} + e^{-\frac{(5-6)^2}{18}} + e^{-\frac{(7-6)^2}{18}} + e^{-\frac{(8-6)^2}{18}}\right] \\ &= 0.1031854 \end{aligned}$$

(c)
- For $h = 1$, log likelihood of validation set $= \log p_1(x = 4) + \log p_1(x = 6) = \log 0.0751318 + \log 0.134517 = -4.59458$

- For $h = 2$, log likelihood of validation set $= \log p_2(x = 4) + \log p_2(x = 6) = \log 0.0971931 + \log 0.1250115 = -4.41041$

- For $h = 3$, log likelihood of validation set $= \log p_3(x = 4) + \log p_3(x = 6) = \log 0.0919010 + \log 0.1031854 = -4.6583$

Since $h = 2$ results in the highest log likelihood in the validation set, therefore $h = 2$ is the best (out of the possibilities of 1, 2 and 3). In practice, one would perform a more exhaustive search over values of h.

2. 2 closest neighbors for $x = 3$ are the sample points 2 and 5. In this case, the volume $V = 2 \times (5 - 3) = 4$. Therefore $p(x = 3) = \frac{2/4}{4} = 1/8$

3. (a) For $x = \begin{pmatrix} 0.5 \\ 1 \end{pmatrix}$, the 3 closest neighbors are x_1, x_2, x_4. Therefore $p(y = 0|x) = 2/3$, while $p(y = 1|x) = 1/3$. Therefore the MAP classification of the class label of x is 0.

   (b) For $x = \begin{pmatrix} -0.5 \\ 1 \end{pmatrix}$, the 3 closest neighbors are x_2, x_3, x_4. Therefore $p(y = 1|x) = 2/3$, while $p(y = 0|x) = 1/3$. Therefore the MAP classification of the class label of x is 1.

   (c) Let $x = \begin{pmatrix} x^1 \\ x^2 \end{pmatrix}$. Then $x^1 = 0$ is the decision boundary. Suppose $x^1 < 0$, then the MAP classification of the class label is 1. Suppose $x^1 > 0$, then the MAP classification of the class label is 0. For $x^1 = 0$, then $p(y = 0|x) = p(y = 1|x) = 1/2$