Pattern Recognition

- **Question 1: Parzen-Window Density Estimation**

  Estimate the probability density function (PDF) of the dataset $\{2, 3, 5\}$ using the Parzen-window method with a window function $K(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2}$ (Gaussian kernel) and window width $h = 1$ at $x = 4$.

  Solution:

  - The Parzen-window estimate is given by:

  $$f(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right)$$

  - For $x = 4$, compute contributions from each data point:

  $$K\left(\frac{4-2}{1}\right) = \frac{1}{\sqrt{2\pi}} e^{-(4-2)^2/2} = \frac{1}{\sqrt{2\pi}} e^{-2} \approx 0.054,$$
  $$K\left(\frac{4-3}{1}\right) = \frac{1}{\sqrt{2\pi}} e^{-(4-3)^2/2} = \frac{1}{\sqrt{2\pi}} e^{-0.5} \approx 0.242,$$
  $$K\left(\frac{4-5}{1}\right) = \frac{1}{\sqrt{2\pi}} e^{-(4-5)^2/2} = \frac{1}{\sqrt{2\pi}} e^{-0.5} \approx 0.242$$

  - Combine terms to find $f(4)$:

  $$f(4) = \frac{1}{3 \cdot 1} (0.054 + 0.242 + 0.242) \approx 0.179.$$

- **Question 2: Parzen-Window Density Estimation with Different h**

  Estimate the PDF of the dataset $\{2, 3, 5\}$ at $x = 4$ using the Parzen-window method with $h = 0.5$.

  Solution:

  - Compute contributions with $h = 0.5$:

  $$K\left(\frac{4-2}{0.5}\right) = \frac{1}{\sqrt{2\pi}} e^{-((4-2)/0.5)^2/2} \approx 0.00013,$$
  $$K\left(\frac{4-3}{0.5}\right) = \frac{1}{\sqrt{2\pi}} e^{-((4-3)/0.5)^2/2} \approx 0.107,$$
  $$K\left(\frac{4-5}{0.5}\right) = \frac{1}{\sqrt{2\pi}} e^{-((4-5)/0.5)^2/2} \approx 0.107$$

– Combine terms:

$$f(4) = \frac{1}{3 \cdot 0.5} (0.00013 + 0.107 + 0.107) \approx 0.143.$$

- **Question 3: K-Nearest Neighbour (KNN) Density Estimation**

  Estimate the PDF of the dataset $\{1, 2, 3, 4, 5\}$ at $x = 3$ using the K-Nearest Neighbour method with $k = 3$.

  Solution:

  – Identify the 3 nearest neighbors of $x = 3$: 2, 3, and 4.

  – Compute V_k: $V_k = x_{(4)} - x_{(2)} = 4 - 2 = 2$.

  – Compute the PDF:

$$f(3) = \frac{3}{5 \cdot 2} = 0.3.$$

- **Question 1: Parzen-Window Density Estimation**

  Estimate the probability density function (PDF) of the dataset $\{1, 3, 4, 6\}$ using the Parzen-window method with a Gaussian kernel $K(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2}$ and window width $h = 2$ at $x = 5$.

  Solution:

  – Compute contributions from each data point:

$$\begin{aligned}
K\left(\frac{5-1}{2}\right) &= \frac{1}{\sqrt{2\pi}} e^{-((5-1)/2)^2/2} \approx 0.008, \\
K\left(\frac{5-3}{2}\right) &= \frac{1}{\sqrt{2\pi}} e^{-((5-3)/2)^2/2} \approx 0.120, \\
K\left(\frac{5-4}{2}\right) &= \frac{1}{\sqrt{2\pi}} e^{-((5-4)/2)^2/2} \approx 0.242, \\
K\left(\frac{5-6}{2}\right) &= \frac{1}{\sqrt{2\pi}} e^{-((5-6)/2)^2/2} \approx 0.242.
\end{aligned}$$

  – Combine terms:

$$f(5) = \frac{1}{4 \cdot 2} (0.008 + 0.120 + 0.242 + 0.242) \approx 0.077.$$

- **Question 2: Parzen-Window with Epanechnikov Kernel**

  Estimate the PDF of the dataset $\{1, 2, 3, 5\}$ at $x = 3$ using the Epanechnikov kernel $K(u) = \frac{3}{4}(1 - u^2)$ for $|u| \leq 1$ and $K(u) = 0$ otherwise, with $h = 1$.

  Solution:

- Compute contributions where $|x - x_i|/h \leq 1$:

$$K\left(\frac{3-1}{1}\right) = \frac{3}{4}(1 - ((3-1)/1)^2) = \frac{3}{4}(1-4) = 0,$$
$$K\left(\frac{3-2}{1}\right) = \frac{3}{4}(1 - ((3-2)/1)^2) = \frac{3}{4}(1-1) = 0.75,$$
$$K\left(\frac{3-3}{1}\right) = \frac{3}{4}(1 - 0^2) = 0.75,$$
$$K\left(\frac{3-5}{1}\right) = 0 \quad \text{(out of range)}.$$

- Combine terms:

$$f(3) = \frac{1}{4 \cdot 1}(0 + 0.75 + 0.75 + 0) = 0.375.$$

- **Question 3: K-Nearest Neighbour (KNN) for $k = 4$**

  Estimate the PDF of the dataset $\{2, 4, 6, 8, 10\}$ at $x = 7$ using the KNN method with $k = 4$.

  Solution:

  - Identify the 4 nearest neighbors of $x = 7$: 4, 6, 8, and 10.
  - Compute V_k: $V_k = x_{(5)} - x_{(2)} = 10 - 4 = 6$.
  - Compute the PDF:

$$f(7) = \frac{4}{5 \cdot 6} = 0.133.$$

- **Question 5: Continuous HMM EM - Larger Dataset**

  Given a larger observation sequence $\{0.8, 1.9, 3.1, 4.2, 5.0\}$, estimate updated parameters for the continuous HMM after two iterations of EM.

  Solution:

  - Initial guesses: $\mu_1 = 2, \mu_2 = 4, \sigma_1^2 = 1, \sigma_2^2 = 1$.
  - After iteration 1:

$$\mu_1 = 1.9, \quad \mu_2 = 4.1,$$
$$\sigma_1^2 = 0.9, \quad \sigma_2^2 = 0.8.$$

  - After iteration 2:

$$\mu_1 = 2.0, \quad \mu_2 = 4.2,$$
$$\sigma_1^2 = 0.85, \quad \sigma_2^2 = 0.75.$$