Problem 1

Question: Estimate the density at $x = 2.5$ for data points $1, 2, 3, 4$ using a Gaussian kernel with bandwidth $h = 1$.

Solution: The Gaussian kernel is given by:

$$K(x, x_i) = \frac{1}{\sqrt{2\pi}h} \exp\left(-\frac{(x - x_i)^2}{2h^2}\right).$$

The density estimate is:

$$f(2.5) = \frac{1}{4} \sum_{i=1}^{4} K(2.5, x_i).$$

Substituting values for each x_i:

$$K(2.5, 1) = \frac{1}{\sqrt{2\pi}(1)} \exp\left(-\frac{(2.5 - 1)^2}{2(1)^2}\right) = \frac{1}{\sqrt{2\pi}} \exp(-1.125) \approx 0.1295,$$

$$K(2.5, 2) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(2.5 - 2)^2}{2}\right) = \frac{1}{\sqrt{2\pi}} \exp(-0.125) \approx 0.3521,$$

$$K(2.5, 3) = \frac{1}{\sqrt{2\pi}} \exp(-0.125) \approx 0.3521,$$

$$K(2.5, 4) = \frac{1}{\sqrt{2\pi}} \exp(-1.125) \approx 0.1295.$$

Now compute:

$$f(2.5) = \frac{1}{4}(0.1295 + 0.3521 + 0.3521 + 0.1295) = \frac{0.9632}{4} \approx 0.2408.$$

Answer: $f(2.5) \approx 0.2408$

—

Problem 2

Question: Find the Parzen window estimate at $x = 1.5$ for data points $1, 2, 2.5, 3$, assuming a rectangular kernel with bandwidth $h = 1$.

Solution: The rectangular kernel is given by:

$$K(x, x_i) = \begin{cases} \frac{1}{2h}, & \text{if } |x - x_i| \leq h, \\ 0, & \text{otherwise.} \end{cases}$$

The density estimate is:

$$f(1.5) = \frac{1}{4}\sum_{i=1}^{4} K(1.5, x_i).$$

For each x_i:

$$K(1.5,1) = \frac{1}{2(1)} = 0.5, \quad K(1.5,2) = 0.5, \quad K(1.5,2.5) = 0.5, \quad K(1.5,3) = 0.$$

Now compute:

$$f(1.5) = \frac{1}{4}(0.5 + 0.5 + 0.5 + 0) = \frac{1.5}{4} = 0.375.$$

Answer: $f(1.5) = 0.375$

—

Problem 3

Question: Compute the probability density estimate at $x = 3.5$ for data $(3, 3, 4, 5)$ using a triangular kernel with $h = 2$.

Solution: The triangular kernel is given by:

$$K(x, x_i) = \begin{cases} \frac{1}{h}\left(1 - \frac{|x-x_i|}{h}\right), & \text{if } |x - x_i| \leq h, \\ 0, & \text{otherwise.} \end{cases}$$

The density estimate is:

$$f(3.5) = \frac{1}{4}\sum_{i=1}^{4} K(3.5, x_i).$$

For each x_i:

$$K(3.5,3) = \frac{1}{2}\left(1 - \frac{|3.5-3|}{2}\right) = \frac{1}{2}(1 - 0.25) = 0.375,$$

$$K(3.5,4) = \frac{1}{2}(1 - 0.25) = 0.375, \quad K(3.5,5) = \frac{1}{2}(1 - 0.75) = 0.125.$$

Now compute:

$$f(3.5) = \frac{1}{4}(0.375 + 0.375 + 0.375 + 0.125) = \frac{1.25}{4} = 0.3125.$$

Answer: $f(3.5) \approx 0.3125$

—

Problem 4

Question: Use a Gaussian kernel with $h = 1$ to find the PDF at $x = 0.5$ for data points $-1, 0, 1, 2$.

Solution: The Gaussian kernel is given by:

$$K(x, x_i) = \frac{1}{\sqrt{2\pi}h} \exp\left(-\frac{(x - x_i)^2}{2h^2}\right).$$

The density estimate is:

$$f(0.5) = \frac{1}{4} \sum_{i=1}^{4} K(0.5, x_i).$$

For each x_i:

$$K(0.5, -1) \approx 0.1295, \quad K(0.5, 0) \approx 0.3521, \quad K(0.5, 1) \approx 0.3521, \quad K(0.5, 2) \approx 0.1295.$$

Now compute:

$$f(0.5) = \frac{1}{4}(0.1295 + 0.3521 + 0.3521 + 0.1295) = 0.2408.$$

Answer: $f(0.5) \approx 0.2408$

—

Problem 5

Question: Estimate the density at $x = 5.5$ with points $(5, 6, 7)$ using an Epanechnikov kernel and $h = 1$.

Solution: The Epanechnikov kernel is given by:

$$K(x, x_i) = \begin{cases} \frac{3}{4h}\left(1 - \left(\frac{x - x_i}{h}\right)^2\right), & \text{if } |x - x_i| \leq h, \\ 0, & \text{otherwise.} \end{cases}$$

The density estimate is:

$$f(5.5) = \frac{1}{3} \sum_{i=1}^{3} K(5.5, x_i).$$

For each x_i:

$$K(5.5, 5) = 0.5625, \quad K(5.5, 6) = 0.5625, \quad K(5.5, 7) = 0.$$

Now compute:

$$f(5.5) = \frac{1}{3}(0.5625 + 0.5625 + 0) = 0.375.$$

Answer: $f(5.5) = 0.375$

Problem 1

Suppose you have 1D data points: $x = \{1, 2, 3, 10, 11, 12\}$. Fit a GMM with $k = 2$ clusters initialized with means $\mu_1 = 2$ and $\mu_2 = 11$. Assume equal variances ($\sigma^2 = 1$) and equal weights. Compute the posterior probability for the first data point ($x_1 = 1$) for each cluster.

Solution

The posterior probability is computed using Bayes' theorem:

$$P(z = j \mid x_i) = \frac{\pi_j \cdot \mathcal{N}(x_i \mid \mu_j, \sigma^2)}{\sum_{k=1}^{K} \pi_k \cdot \mathcal{N}(x_i \mid \mu_k, \sigma^2)}$$

Step 1: Compute the likelihood ($\mathcal{N}(x_i \mid \mu_j, \sigma^2)$) for both clusters. For $x_1 = 1$:

$$\mathcal{N}(1 \mid 2, 1) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(1-2)^2}{2 \cdot 1}\right) = 0.24197$$

$$\mathcal{N}(1 \mid 11, 1) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(1-11)^2}{2 \cdot 1}\right) = 7.998 \times 10^{-23}$$

Step 2: Compute posterior probabilities. Assume equal weights ($\pi_1 = \pi_2 = 0.5$):

$$P(z = 1 \mid x_1) = \frac{0.5 \cdot 0.24197}{0.5 \cdot 0.24197 + 0.5 \cdot 7.998 \times 10^{-23}} = 1$$

$$P(z = 2 \mid x_1) = \frac{0.5 \cdot 7.998 \times 10^{-23}}{0.5 \cdot 0.24197 + 0.5 \cdot 7.998 \times 10^{-23}} = 0$$

Answer: $P(z = 1 \mid x_1) = 1, P(z = 2 \mid x_1) = 0$.

Problem 2

Given the data points $x = \{2, 4, 6, 8\}$, initialize a GMM with two clusters ($k = 2$) with means $\mu_1 = 3$, $\mu_2 = 7$, variances $\sigma_1^2 = \sigma_2^2 = 1$, and equal weights. Compute the E-step responsibilities for $x_2 = 4$.

Solution

Likelihood:

$$\mathcal{N}(4 \mid \mu_1 = 3, \sigma_1^2 = 1) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(4-3)^2}{2 \cdot 1}\right) = 0.24197$$

$$\mathcal{N}(4 \mid \mu_2 = 7, \sigma_2^2 = 1) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(4-7)^2}{2 \cdot 1}\right) = 0.00443$$

Responsibilities:

$$R_{1,2} = \frac{\pi_1 \cdot \mathcal{N}(4 \mid 3, 1)}{\pi_1 \cdot \mathcal{N}(4 \mid 3, 1) + \pi_2 \cdot \mathcal{N}(4 \mid 7, 1)}$$

Assume equal weights ($\pi_1 = \pi_2 = 0.5$):

$$R_{1,2} = \frac{0.5 \cdot 0.24197}{0.5 \cdot 0.24197 + 0.5 \cdot 0.00443} = 0.982$$

$$R_{2,2} = \frac{0.5 \cdot 0.00443}{0.5 \cdot 0.24197 + 0.5 \cdot 0.00443} = 0.018$$

Answer: $R_{1,2} = 0.982, R_{2,2} = 0.018.$

Problem 3

The same data points as Problem 2. Compute updated cluster weights after one E-step.

Solution

Cluster weights are updated as:

$$\pi_j = \frac{1}{N} \sum_{i=1}^{N} R_{j,i}$$

For $k = 2$, $x = \{2, 4, 6, 8\}$: Compute $R_{j,i}$ as in Problem 2 for all points, then sum and normalize to update the weights.

Problem 1

Consider a dataset of coin flips with two biased coins A and B. You observe the following sequence of outcomes: $\{H, H, T, H, T\}$. Coin A has $P(H|A) = 0.6$ and $P(T|A) = 0.4$, while coin B has $P(H|B) = 0.3$ and $P(T|B) = 0.7$. Assume an initial probability $\pi_A = 0.5$ and $\pi_B = 0.5$. Compute the expected likelihood of each coin generating the first outcome (H) during the E-step.

Solution

The expected likelihood is computed using Bayes' theorem:

$$P(A \mid H) = \frac{\pi_A \cdot P(H \mid A)}{\pi_A \cdot P(H \mid A) + \pi_B \cdot P(H \mid B)}$$

Step 1: Calculate likelihoods.

$$P(H \mid A) = 0.6, \quad P(H \mid B) = 0.3$$

Step 2: Compute posteriors.

$$P(A \mid H) = \frac{0.5 \cdot 0.6}{0.5 \cdot 0.6 + 0.5 \cdot 0.3} = \frac{0.3}{0.3 + 0.15} = 0.6667$$

$$P(B \mid H) = \frac{0.5 \cdot 0.3}{0.5 \cdot 0.6 + 0.5 \cdot 0.3} = \frac{0.15}{0.3 + 0.15} = 0.3333$$

Answer: $P(A \mid H) = 0.6667, P(B \mid H) = 0.3333.$

—

Problem 2

Using the same setup as Problem 1, calculate the expected number of heads generated by each coin if the entire sequence is $\{H, H, T, H, T\}$.

Solution

For each observation, compute the responsibility $(R_{i,j})$ for each coin j:

$$R_{i,A} = P(A \mid x_i), \quad R_{i,B} = P(B \mid x_i)$$

Step 1: Compute responsibilities. For H:

$$R_{H,A} = 0.6667, \quad R_{H,B} = 0.3333$$

For T, the posterior probabilities are:

$$P(A \mid T) = \frac{\pi_A \cdot P(T \mid A)}{\pi_A \cdot P(T \mid A) + \pi_B \cdot P(T \mid B)}$$

$$P(A \mid T) = \frac{0.5 \cdot 0.4}{0.5 \cdot 0.4 + 0.5 \cdot 0.7} = \frac{0.2}{0.2 + 0.35} = 0.3636$$

$$P(B \mid T) = 0.6364$$

Step 2: Compute expected counts. The total expected counts of heads generated by A and B are:

$$\text{Heads by } A = \sum_{i \in H} R_{i,A} = 0.6667 + 0.6667 + 0.6667 = 2.0$$

$$\text{Heads by } B = \sum_{i \in H} R_{i,B} = 0.3333 + 0.3333 + 0.3333 = 1.0$$

Answer: Expected heads by $A = 2.0$, Expected heads by $B = 1.0$.

—

Problem 3

Continue from Problem 2. Update the probabilities π_A and π_B after the E-step.

Solution

Step 1: Compute updated probabilities. The total number of flips is $N = 5$. The updated probabilities are:

$$\pi_A = \frac{\sum_{i=1}^{N} R_{i,A}}{N}, \quad \pi_B = \frac{\sum_{i=1}^{N} R_{i,B}}{N}$$

$$\pi_A = \frac{2.0 + 0.7272}{5} = 0.5454$$
$$\pi_B = \frac{1.0 + 0.7272}{5} = 0.4546$$

Answer: $\pi_A = 0.5454, \pi_B = 0.4546$.

—

Problem 4

Using updated probabilities from Problem 3, calculate the new likelihood for the dataset during the M-step.

Solution

The likelihood for a dataset is:

$$\mathcal{L} = \prod_{i=1}^{N} (\pi_A \cdot P(x_i \mid A) + \pi_B \cdot P(x_i \mid B))$$

Step 1: Compute for each observation. For H:

$$P(H) = 0.5454 \cdot 0.6 + 0.4546 \cdot 0.3 = 0.4458$$

For T:

$$P(T) = 0.5454 \cdot 0.4 + 0.4546 \cdot 0.7 = 0.5542$$

Step 2: Multiply probabilities.

$$\mathcal{L} = (0.4458)^3 \cdot (0.5542)^2 = 0.03068$$

Answer: Likelihood $\mathcal{L} = 0.03068$.

—

Problem 5

Verify convergence by comparing the updated and previous parameters. Use the results from Problems 3 and 4 to determine if the parameters have stabilized.

Solution

Convergence is typically determined by comparing π_A, π_B and likelihood values across iterations. If changes are below a set threshold (ϵ), the algorithm has converged.

Problem 1

Given the following dataset of points in 2D space:

$$\{(1,2),(2,3),(3,3),(6,6),(7,7)\}$$

Perform a single-linkage hierarchical clustering algorithm. Compute the initial distance matrix and show the first two merging steps.

Solution

Step 1: Compute the distance matrix. The Euclidean distance between two points (x_1, y_1) and (x_2, y_2) is:

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

The initial distances between points are:

$$\begin{bmatrix} & (1,2) & (2,3) & (3,3) & (6,6) & (7,7) \\ (1,2) & 0 & 1.41 & 2.24 & 6.40 & 7.62 \\ (2,3) & 1.41 & 0 & 1 & 5 & 6.40 \\ (3,3) & 2.24 & 1 & 0 & 4.24 & 5.66 \\ (6,6) & 6.40 & 5 & 4.24 & 0 & 1.41 \\ (7,7) & 7.62 & 6.40 & 5.66 & 1.41 & 0 \end{bmatrix}$$

Step 2: Merge the closest points. - The closest points are $(2,3)$ and $(3,3)$, with $d = 1$. Merge these into a single cluster $\{(2,3),(3,3)\}$. - Update the distance matrix by taking the minimum distance (single linkage) between the new cluster and all other points.

$$\begin{bmatrix} & (1,2) & \{(2,3),(3,3)\} & (6,6) & (7,7) \\ (1,2) & 0 & 1.41 & 6.40 & 7.62 \\ \{(2,3),(3,3)\} & 1.41 & 0 & 4.24 & 5.66 \\ (6,6) & 6.40 & 4.24 & 0 & 1.41 \\ (7,7) & 7.62 & 5.66 & 1.41 & 0 \end{bmatrix}$$

Step 3: Merge the next closest points. - The closest points are $(6, 6)$ and $(7, 7)$, with $d = 1.41$. Merge these into a cluster $\{(6, 6), (7, 7)\}$.

Continue this process until all points are in a single cluster.

—

Problem 2

Using the dataset from Problem 1, construct a dendrogram to visualize the clustering process.

Solution

The dendrogram represents the merging of clusters based on the distance metric. Each merge corresponds to a horizontal line at the height of the distance where the merge occurs.

Draw the dendrogram manually based on the steps in Problem 1.

—

Problem 3

Describe how the clustering result would differ if complete linkage was used instead of single linkage.

Solution

Complete linkage uses the maximum distance between points in two clusters to determine the distance. This generally results in more compact clusters but can delay merging larger groups. Recompute the distance matrix using the maximum distance at each step.

—

Problem 1

Given the dataset:

$$\{(1, 1, \text{Class A}), (2, 2, \text{Class A}), (3, 3, \text{Class B}), (6, 6, \text{Class B})\}$$

Predict the class of the point $(4, 4)$ using $k = 3$.

Solution

Step 1: Compute distances to all points. Using Euclidean distance:

$$d((4, 4), (1, 1)) = \sqrt{(4-1)^2 + (4-1)^2} = 4.24$$

$$d((4, 4), (2, 2)) = \sqrt{(4-2)^2 + (4-2)^2} = 2.83$$

$$d((4, 4), (3, 3)) = \sqrt{(4-3)^2 + (4-3)^2} = 1.41$$

$$d((4, 4), (6, 6)) = \sqrt{(4-6)^2 + (4-6)^2} = 2.83$$

Step 2: Select the 3 nearest neighbors. The 3 nearest neighbors are:

$$(3, 3, \text{Class B}), (2, 2, \text{Class A}), (6, 6, \text{Class B})$$

Step 3: Determine the majority class. Class B occurs twice, Class A occurs once. Thus, $(4, 4)$ is predicted to belong to **Class B**.

Answer: $(4, 4) \rightarrow$ Class B.

—

Problem 2

For the dataset in Problem 1, compute the accuracy of the KNN algorithm if $k = 1$ and the test points are:

$$\{(1.5, 1.5, \text{Class A}), (5, 5, \text{Class B})\}.$$

Solution

Step 1: Predict the class for $(1.5, 1.5)$ with $k = 1$. Nearest neighbor: $(1, 1, \text{Class A})$. Predicted class: **Class A**.

Step 2: Predict the class for $(5, 5)$ with $k = 1$. Nearest neighbor: $(6, 6, \text{Class B})$. Predicted class: **Class B**.

Step 3: Compute accuracy. Both predictions are correct. Accuracy:

$$\text{Accuracy} = \frac{\text{Correct Predictions}}{\text{Total Predictions}} = \frac{2}{2} = 1.0$$

Answer: Accuracy = 100%.

—

Problem 3

Explain how the choice of k affects the performance of KNN.

Solution

- **Small k:** Sensitive to noise, leading to overfitting. - **Large k:** Smoothens the decision boundary but may misclassify points if the dataset has imbalanced classes.

—