Teeranat Yingpunyachok
BlazerID: TYINGPUN

Question 1

1. Pseudo-code

Function KNN(X, x , k):

Inputs:
X: List of data point x
    x: given data point
    k: k nearest neighbors

Steps:
    Initial **Distances** as an Empty List

    For each **Item** in X:

        If Item != x:

            Compute **distance** between x and Item

            Append (**Item**, **distance**) into **Distances**

    Sort **distances** in ascending order by **distance**

    Select the first **k** elements from **Distances**

    Return **distances**

Output:

Return the list of k nearest neighbors from the List of data point X to the given data point x

2. Pseudo-code
Function KNN_detect_outliers(X, k, d_h):
Inputs:
    X: List of data point x
    k: k nearest neighbors
    D_h: detection threshold
Steps:
1. Initial
Initial **outlier_points** as an Empty List
    2. calculate each point in X
    For each point **xi** in **X**:

2.1 call KNN(X,xi,k) > return **k-neighbors** list
2.2 compute outlier_score = 1/k * (sum of **k-neighbors**)
2.3 if **outlier_score** > **D_h**:
Append **xi** into **outliers_points**
Return **outlier_points**

Output:

Outlier_points: list of outlier points

Question 2

Intitial

1. From PDF of Gaussian distribution

$$f(\, x_i \;\mid\; \mu, \sigma^2\,) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right)$$

2. Take log to PDF of Gaussian distribution

$$\log f\,(\, x_i \mid \mu, \sigma^2\,) = -\frac{(x_i-\mu)^2}{2\sigma^2} - \log\left(\sqrt{2\pi\sigma^2}\right)$$

3. From Total MLE

$$L(\theta) = \Pi_{i=1}^{N} f(\, x_i \mid \theta\,)$$

4. Take log to MLE and $\theta = \mu,\ \sigma^2$

$$\log\big(L(\mu,\sigma^2)\big) = \log\left(\Pi_{i=1}^{N} f(\, x_i \mid \mu, \sigma^2\,)\right)$$

$$\log\big(L(\mu,\sigma^2)\big) = \sum_{i=1}^{n} \log f\,(\, x_i \mid \mu, \sigma^2\,)$$

5. Substitute from #2 to #4

$$\log\big(L(\mu,\sigma^2)\big) = \sum_{i=1}^{n}\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right) - \frac{n}{2}\left(\log\left(\sqrt{2\pi\sigma^2}\right)\right)$$

Find μ , Take a derivative by μ

$$\frac{\partial}{\partial\mu}\log(L) \;=\; \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)$$

Because we find Maximum Likelihood estimation, so we make derivative = 0 so that:

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu) = 0\,;\ \sigma\ > 0$$

Teeranat Yingpunyachok
BlazerID: TYINGPUN

$$\sum_{i=1}^{n}(x_i - \mu) = 0$$

$$\sum_{i=1}^{n}(x_i) - n\mu = 0$$

$$\mu = \frac{1}{n}\sum_{i=1}^{n}(x_i)$$

Find σ^2, Take derivative by σ^2

$$\frac{\partial}{\partial\sigma^2}\log(L) = \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i - \mu)^2 - \frac{n}{2\sigma^2}$$

Because we find Maximum Likelihood estimation, so we make derivative = 0 so that:

$$\frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i - \mu)^2 - \frac{n}{2\sigma^2} = 0$$

$$\frac{n}{2\sigma^2} = \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i - \mu)^2 \ ; \ n > 0, \ \sigma^2 > 0$$

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$$