Question 1:

DONE IN NOTEBOOK

Question 2:

a) As most Machine Learning models face similar complications, having an incorrect number of clusters can lead to overfitting and underfitting. Assume that our model has a significantly lesser number of clusters than required, this means that the model will fail to understand the differences between certain points implying that it is too simple to capture and interpret our data. Now let us assume the other extreme where our model has too many clusters than required, this means that our model is extremely sensitive to noise and outliers which may cause a significant number of problems due to our model being too complex. Choosing an incorrect number of clusters will certainly always provide us with skewed results which can be extremely devastating when deployed.

b) Choosing an incorrect number of clusters would cause many implications and one of the most significant ones is having incorrect predictions and results. If we take the practicality stand point, having our model exhibit such predictions can cause devastating damages. Let's say our AI model is placed in the medical field, this means that any incorrect results can cause severe and extreme damages to the patients.

Question 3:

DONE IN NOTEBOOK

Question 4:

a) Outliers can simply be defined as data points that tend to have very little correlation with the majority of other points in terms of interpreting what the data points signify.

b) Outlier detection can allow us to identify outliers which allows us to either delete them for the purpose of minimizing noise or we can choose to either keep and use them as information. (Fraud detection etc.)

c) Local outlier detection is a method that is used to identify outliers by comparing the density of a given point to the density of its neighboring points. On using this method, we can consider a point to be an outlier if the density of the point is much lower than that of its neighbors.

Question 5:

a) The process of Nonparametric density using histograms is as follows:

   i) Divide the input space into bins of equal size.

   ii) From a chosen origin x_0 and bin width h, and the created intervals, the estimate is equal to *sum of all data points / bin width * number of data points.*

The advantage of using this method comes down to its simplicity and ease of interpretability. Nonparametric Density Estimation also does a great job in the identification of outliers.

b) The kernel estimator sums the bumps. The bumps are all the x^t values that have an effect on the estimate of x. These bumps can be categorized by a Guassian function. The effect of this estimator will cause the estimate to eventually produce a graph which gets noticeably smoother.
This differs from the knn-estimator approach as the kernel estimator uses all the data points in the training set for the classification of new data points whereas the k-nn approach uses only *k* number of points that are the closest. The kernel as we have discussed above always provides a smoother estimate whereas the k-nn estimator tends to have more linear and spikey boundaries in comparison to the kernel estimator.

Question 6

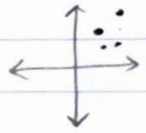
a) DONE IN NOTE BOOK

b) The Gaussian kernel is more isotropic in nature, meaning that it contributes equally in all directions. This results in a spherical shape in the density estimation. Inversely, The Ellipsoidal kernel, in contrast, can be anisotropic, allowing for density contributions that are not equal in all directions.
The Gaussian Kernel also uses the euclidean distance function whereas the Ellipsoidal Kernel uses the Mahalanobis Distance function.

c) A CNN algorithm is an algorithm where the classification is performed by a 1-nn. The algorithm starts with an empty Z class (condensed set) and passes over the instance of elements in the set X in random order. As it iterates through the instances, if the instance is misclassified, it is added to the condensed set and will be considered as new information that has not been captured by the set. The algorithm stops when the sum of all instances either stops or converges towards one value and the Z (condensed set) will now become our consistent set of our original data set.

Question 1 : $A = (1,1)$ $B = (2,2)$ $C = (2,4)$ $D = (1,2)$

a)



- Step 1 : Each point as its own cluster

dist(A, B) : $\sqrt{(2-1)^2 + (2-1)^2} = \sqrt{2}$

dist(A, C) : $\sqrt{10}$

dist(A, D) : (1) New cluster = (A + B)

dist(B, C) : 2

dist(B, D) : 1

dist(C, D) : $\sqrt{5}$

- Step 2 : Consider A · D as cluster

dist(AD, B) $\Rightarrow \sqrt{(1-2)^2 + (1-2)^2} = \sqrt{2}$

(OR) $\Rightarrow \sqrt{(1-2)^2 + (2-2)^2} = 1$ • Cluster (AD + B)

dist(AD, C) $\Rightarrow \sqrt{(1-2)^2 + (1-4)^2} = \sqrt{10}$

(OR) $\Rightarrow \sqrt{(1-2)^2 + (2-4)^2} = \sqrt{5}$ •

dist(B, C) $\Rightarrow \sqrt{(2-2)^2 + (2-4)^2} = 2$

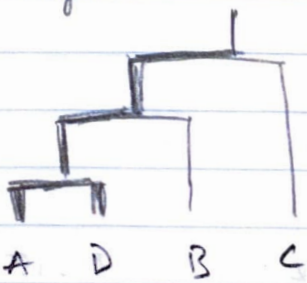
- Step 3 : Consider A · D · B as cluster

dist(ADB, C) $\Rightarrow \sqrt{(1-2)^2 + (1-4)^2} = \sqrt{10}$

(OR) $\Rightarrow \sqrt{(1-2)^2 + (2-4)^2} = \sqrt{5}$

(OR) $\Rightarrow \sqrt{(2-2)^2 + (2-4)^2} = 2$ • Cluster ADBC

b) Dendogram :



Question 3

data points = [2, 3, 5, 7, 9]

a) $x = 4$: $|4 - 2| = 2 : d_1$

$|4 - 3| = 1 : d_2$ •

$|4 - 5| = 1 : d_3$

$|4 - 7| = 3 : d_4$

$|4 - 9| = 5 : d_5$

Sorted distance matrix : $[\underset{d_1}{1}, \underset{d_2}{1}, \underset{d_3}{2}, \underset{d_4}{3}, \underset{d_5}{5}]$

$\Rightarrow x = 4, d_k = d_2(x) = 1, N = 5$ (Sample size)

$$\hat{p}(x) = \frac{k}{2N d_k(x)} = \frac{2}{2 \cdot 5 \cdot 1} = \frac{2}{10} = \frac{1}{5}$$

$\therefore$ density estimation of $x = 4, k = 2 = 0.2$

b) $x = 5, k = 3$

$: |5 - 2| = 3 : d_1$

$|5 - 3| = 2 : d_2$

$|5 - 5| = 0 : d_3$ •

$|5 - 7| = 2 : d_4$

$|5 - 9| = 4 : d_5$

Distance Matrix = $[3, 2, 0, 2, 4]$

Sorted distance matrix = $[\underset{d_1}{0}, \underset{d_2}{2}, \underset{d_3}{2}, \underset{d_4}{3}, \underset{d_5}{4}]$

$\Rightarrow x = 5, d_3(5) = 2, N = 5, k = 3$

$$\Rightarrow \hat{p}(x) = \frac{k}{2 \cdot N \cdot d_k(x)} = \frac{3}{2 \cdot 5 \cdot 2} = \frac{3}{20}$$

$\therefore$ density estimation of $x = 5, k = 3 = 0.15$

Question 6 :

a) bin size = 0.5 , N(Sample Space) = 10

when $x = 3$

$$\hat{p}(x) = \frac{1}{10 \cdot 0.5} = \frac{1}{5} = 0.2$$

$$\hat{p}(3) = \frac{\#\ x^t \text{ in bin } x_3}{N \cdot h}$$

Since 3.1 is the only point in the which is the present in the same bin ,

$\therefore$ There is only one point