Homework -5

Ques 1:

1. Reducing the number of features in a dataset, known as dimensionally reduction, can be achived using techniques such as PCA. For Instance Based learning such as k-Nearest Neighbours (KNN), dimmensionality reduction is important as it eliminates noisy or redundant features, which can negatively affect KNN's performance. Irrelevant features can also ~~improve~~ increase computational complexity making KNN computationally expensive and slow. By reducing the number of dimensions, the "curse of dimensionality" problem can be addressed, allowing KNN to work more effectively with training instances.

2. k-Means is a popular clustering algorithm that suffers from the issue of variability where its results can be greatly affected by the initial placement of cluster centroids. To overcome this limitation, several approaches can be taken.

One method is to run k-Means multiple times with different random initializations, and then select the best clustering solution based on a predefined criterion. This is known as k-Means ++ initialization, where the algorithm starts with one ~~centr~~ centroid randomly chosen from the data points and then selects subsequent centroids by choosing the farthest data points from the existing centroids. This approach increases the chances of finding a good clustering solution by spreading the centroids in the feature space.

Another way to reduce the variability in k-Means is to use a variant called k-Means with mini-batches, where the algorithm updates the centroids using a random subset of data points at each iteration.

3. Gaussian Mixture Model is a ~~statist~~ statistical model that represents a dataset as a mixture of multiple Gaussian distributions. This method is widely used for detecting anomalies

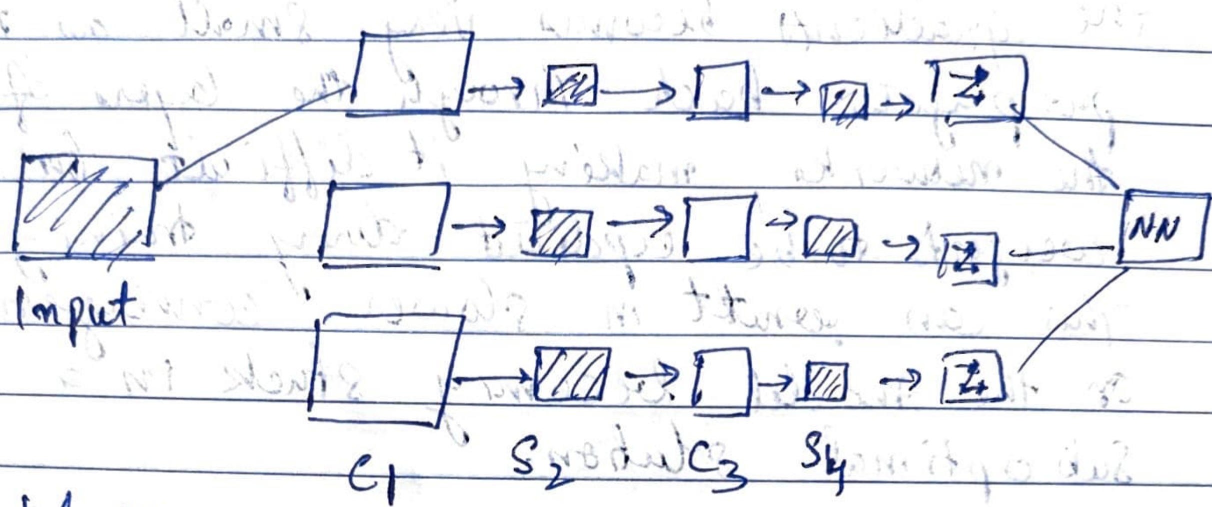
which are data points that devices significant from normal behavior of the majority of data point that ~~dove~~ deviate significantly from the normal behavior of the majority of data points

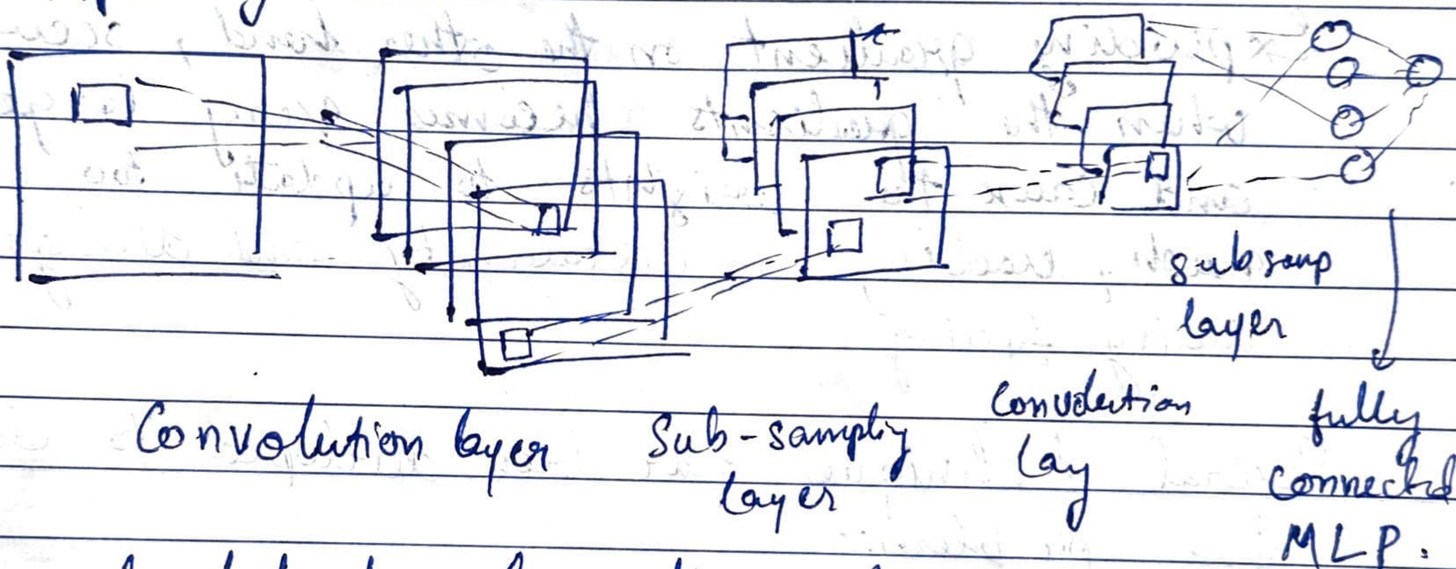
To detect anomalies using GMM, the first step is to train the model on a dataset containing only normal data points. The GMM algorithm then estimates the mean and covariance of each Gaussian distribution based on the training data.

Next, the probability of each data point in the test dataset is calculted using the GMM. Data points that have a high probality of belonging to the normal distrubion are considered normal, while those with low probability are considered anamolous.

A threshold is then set based on the probability distribution of the data points. Any data point that falls below this threshold is considered an anomaly and flagged as a potential outlier. Finally the test dataset is evaluated, and any data point with probability below threshold is identify as an anormality.

4. CNN





Several latest algorithms of CNN are

1. LetNet-5
2. Alex Net
3. GoogLe Net
4. VGG Net
5. Res Net
6. Inception-V4
7. SENet
8. YOLO
9. Capsule Network

5. Vanishing ~~and exploding~~ occurs when the gradients becomes very small as they propagate back through the layers of the networks, making it difficult for the weights to be updated during training. This can result in slower convergence or the model becoming stuck in a suboptimal solution.

Exploding gradient, on the other hand, occur when the gradients becomes very large and cause the weights to update too much, leading to instability and divergence during training.

Several techniques have been developed to address these problem:-

1. Weight initialization.
2. Gradient clipping
3. Batch normalization.
4. Residual connections
5. Learning rate schedule
6. Long-short term memory.

Que 2

Given

$n = 100$

$x = 20$

x is No. of incorrectly classified hypothesis.

$\hat{p} = x/n = \frac{20}{100} = 0.20$

C% confidence interval for population proportion $(\hat{p} - E, \hat{p} + E)$

$$E = z' * \sqrt{\frac{\hat{p} \times (1-\hat{p})}{n}}$$

$$C\% \Rightarrow \left[\left[0.8 - 1.96 * \sqrt{\frac{0.80 \times 2}{100}}\right], 0.8 + 1.96 \sqrt{\left(\frac{0.8 * 2}{100}\right)}\right]$$

$$\Rightarrow [0.707, 0.893]$$

Therefore, we can be 95% confident that the true error rate of hypothesis h on the underlying distribution is between 70.7% & 89.3%.