

Q1) In linear A.M.L. HW - 5 it is asked that if the

- Q1) What will happen if the dimensionality of the dataset tends to increase?
- 1) We all know that with an increase in the dimension, the size of the dataset tends to increase which will lead to an increase in the gap between the datapoints. This causes a decrease in the efficiency of the algorithm.

During such a case use of PCA can help in improving the performance at a very low cost of accuracy. It has benefits such as:

- ① producing independent, uncorrelated features of data
- ② reduction in noise.

Some of the problems that are caused by large data are ① leads to overfitting ② Long computational time

- mismatched elements

- 2) The fundamental issue with k-means variability is that it assumes that every variables distribution variance is spherical. This implies that it is vulnerable to outliers & each of the k-clusters has the prob of the num of observation.

Ways to address the problem -

- Inject some randomness in initialization pts
- Along with random initialization try increasing the no. of repeats for better handling of errors.

3) A GMM model is a probabilistic model in which all data points are generated by combining a finite number of Gaussian distributions with unknown parameters.

~~about frequent soft assignment~~  
 GMM  $\rightarrow$  Generalization of k-means clustering  
 GMM contains info about the data's covariance structure and latent Gaussian centres.

~~using m=1~~  
 $f(x) = \sum_{m=1}^M \phi(x; \mu_m, \Sigma_m)$  where,  
 $\phi(x; \mu_m, \Sigma_m)$  is a Gaussian fn with mean  $= \mu_m$  and covariance matrix  $\Sigma_m$

$$P(x_i \in C_m) = \frac{\phi(x_i; \mu_m; \Sigma_m)}{\sum_{k=1}^M \phi(x_i; \mu_k; \Sigma_k)}$$

Anomaly detection:-

1) Univariate Case - One-D model

$$P(x) = \sum_{i=1}^k \phi_i N(x | \mu_i, \sigma_i^2)$$

$$N(x | \mu_i, \sigma_i^2) = \frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{(x - \mu_i)^2}{2\sigma_i^2}}$$

Here  $\mu_x$  = mean

$\sigma_x^2$  = Variance for  $k^{th}$  element

$\phi_x$  = weight for cluster  $x$

2) For multivariate case : multi Dimensional

$$P(\vec{x}) = \sum_{i=1}^k \phi_i N(\vec{x} | \vec{\mu}_i, \Sigma_i)$$

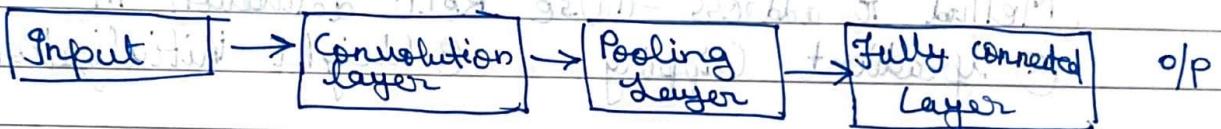
$\vec{\mu}_x$  = mean  $\Sigma_x$  = covariance matrix of  $x^{th}$  element

$\phi_k = \text{weight of cluster } k$  (2)

Here, the algo trains upon the  $k$  clusters. Hence when we receive a new data pt, the algo determines the dist from every distrib' & gives out the prob. of belonging to each cluster.

If we have a cluster with the low prob then it tells us the data pt is an anomaly.

47



Convolution layer - We perform convolution operation in order to extract features of the image.

Activation layer is a sublayer of convolution layer which consists of activation functions like sigmoid, ReLU, etc.

Pooling layer - This is a sampling layer to reduce size of i/p & to try & keep overall value of the i/p same.

Fully connected layer - It is similar to a regular neural network. It is the final learning phase of the network which maps to extracted features of o/p.

CNN algos  $\rightarrow$  AlexNet, ResNet, LeNet

5) Vanishing gradients is the problem of gradient is gradually too small by accumulation and tends to disappear in backpropagation, like the number of gradient is smaller than 1. Exploding gradients is the problem of gradient is gradually too large and by accumulation tend to disappear in Backpropagation like the number of gradient is greater than 1.

Method to address - ① Use ReLU activation fn  
 ② Gradient Clipping      ③ Weight initialization

Q2) i) no. of samples classified incorrectly  
 $= 100 - 20 = 80$   
 ii) sample size = 106  
 $\hat{p}$  - rate for Error,  $p(h)$

$\hat{p} = \frac{20}{100} = 0.2$   
 c.i. confidence interval =  $(\hat{p} - E, \hat{p} + E)$  where

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

margin of error

$z^*$  is critical  $\approx z$  score for a particular confidence interval

$$\alpha = 1 - C = 1 - 0.95 = 0.05$$

$$\frac{\alpha}{2} = 0.025$$

$$z^* = Z_{0.025} = 1.96$$

$$E = 1.96 \sqrt{\frac{0.2(1-0.2)}{100}} = 0.0784$$

q.s.i. = confidence interval for true error rate  
 for  $E_{\text{error}}(h) = (0.2 - 0.0784, 0.2 + 0.0784)$   
 $= (0.1216, 0.2784)$