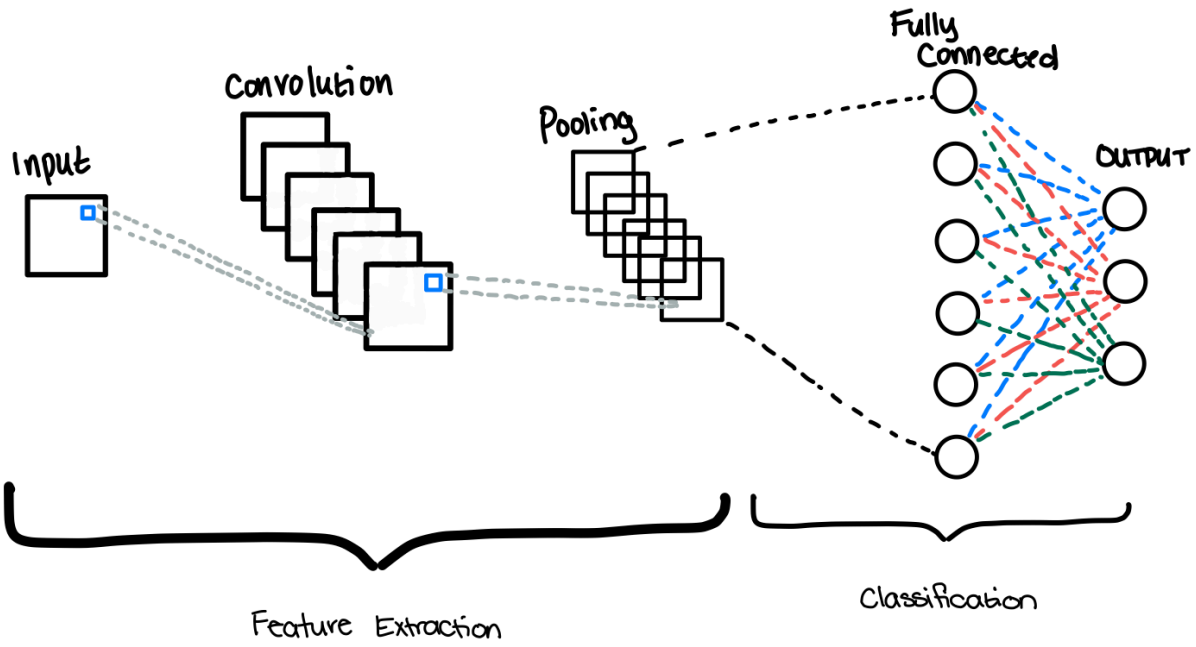
For Questions 1- 2, please submit a word file or a PDF file;
For Question 3 (programming question), please submit an .ipynb file.

Question 1 (15 points):

1) **Please explain the technique of Gaussian Mixture and how it is used for anomaly detection. (5 points)**

Gaussian Mixture uses probability clustering. The model assumes that each data point belongs to a Gaussian distribution and uses anomaly detection to detect outliers. Outliers would be considered the data points found in low-density regions.

2) **Please draw the diagram of Convolutional Neural Networks (CNN). Then explain the functionality of each layer of CNN. Name several latest algorithms of CNN (e.g., AlexNet etc.). (5 points)**



A neural network is a standard network that consists of multiple layers that are interconnected. Each layer receives an input, transforms that input to something else, and passes an output to the next layer.

CNN is a part of the neural network and is a section of layers. In these layers are filters that perform pattern recognition.

Layers

1. Convolution
    a. Places a filter over an array of image pixels, this then creates a convolved feature map.
2. Pooling

    a. Reduces the sample size of a particular feature map, but retains important information.
    b. The output of this is a pooled feature map;
        i. Max pooling: takes the max input of a particular convolved feature
        ii. Average pooling: Takes the average input of a particular convolved feature
        iii. Sum Pooling: Takes the sum input of a particular convolved feature
3. Fully Connected Layer
    a. The matrix is flattened into a vector and fed to the fully connected layer

Latest CNN algorithms

- LeNet-5
- VGG-16
- VGG-19
- Inception and GoogLeNet
- ResNet

3) **What are the vanishing and exploding gradients problems in Backpropagation? Name several techniques to address these problems. (5 points)**

A vanishing gradient is when, in a network of n hidden layers, the multiplied derivatives are small causing the gradient to shrink exponentially. They will eventually become too small and vanish. The results of a model using small gradients are that it will be unable to learn meaningful insights due to the weights and biases of the initial layers not being updated effectively.

An exploding gradient is when, in a network of n hidden layers, the multiplied derivatives are large causing the gradient to increase exponentially until they explode. A model with large weights and biases will be unstable, possibly causing an overflow resulting in NaN weight values that can't be further updated, and unable to effectively learn.

Some techniques to address these problems would be reducing the amount of layers (vanishing, exploding), gradient clipping (exploding), and weight initialization (vanishing/exploding).

Question 2 (5 points):

Consider a learned hypothesis, *h*, for some Boolean concept. When *h* is tested on a set of 100 examples, it classifies 80 correctly. What is the 95% confidence interval for the true error rate for $Error_D(h)$?

Calculation of Standard Deviation Estimate

$$100(0.20)(1-0.20) = 16$$

$$\sqrt{16} = 4$$

$$4/100 = 0.04$$

95% confidence interval for $Error_D(h)$ =

$$Error_D(h) +- 1.96 \cdot \sqrt{[Error_D(h) \cdot (1 - Error_D(h)) / n)}$$

$$= 0.20 +- 1.96 \cdot \sqrt{[0.20(1-0.20)/100]}$$

$$= 0.20 + 0.0784 = 0.2784$$

$$= 0.20 - 0.0784 = 0.1216$$