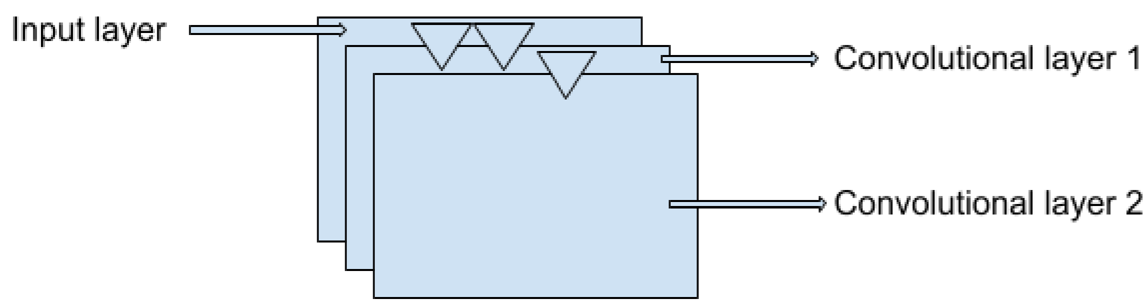
Joscandy Nunez
695 Applied Machine Learning

Question 1 (15 points):
1. Gaussian Mixture is a probabilistic model. The instances from one Gaussian distribution form a cluster. Each of these clusters has their own means and covariance matrices. In a given dataset, one needs to estimate the weights as well as the distribution parameters. To find these, the Expectation-Maximization (EM) algorithm is used.

This algorithm first initializes cluster parameters randomly, and then:
1) Assigns instances to clusters (expectation step)
2) Updates the clusters (maximization step)

These two steps repeat until convergence, and each instance is weighted by the probability that it belongs to that particular cluster.

Using Gaussian Mixture can detect anomalies by identifying outliers of the Gaussian Mixture model. This is done by finding the clusters' centers, size, shape, and orientation ($\Sigma(1)$ to $\Sigma(k)$), and relative weights ($\varphi(1)$ to $\varphi(k)$). The model pairs each instance with its corresponding cluster using the probability distribution and density.



2. The rectangles (layers) are stacked up on top of each other (3D) and the pyramids are touching the base of the rectangle they are on and connecting with the rectangle (layer) on top.

Convolution layer applies learned filters and outputs feature maps.
Pooling layer subsample feature maps (for efficiency).
Learning goes from low-level to high-level features.
Features are normalized between layers.

Latest algorithms : AlexNet, Inception v-1, Xception, LeNet-5, Inception ResNets

3. The vanishing gradients problem happens when gradients get exponentially smaller as the algorithm gets to lower layers. As a result, the Gradient Descent update leaves the lower layers' connection weights almost unchanged, and the training never converges. This does not lead to a good solution.

Exploding gradients problem happens when the gradients get exponentially bigger, and the layers get large weight updates, and the algorithm diverges.

Techniques to address these problems :

Joscandy Nunez
695 Applied Machine Learning

- Batch Normalization
- Gradient Clipping
- Reusing Pretrained Layers or “transfer learning”
- Pretraining on an Auxiliary Task
- Learning Rate Scheduling
- Momentum Optimization