Chapter 2: Mathematical building blocks of neural networks

Deep learning applies several mathematical concepts like Tensor (data, storing object going into the network),Tensor operation(activities that occur in each layer),gradient descent(which allow network to learn from the train samples).these concepts serve as the basic building block of neural networks and their understanding is the key to understanding the activities in a neural network.

Tensors (Data structure of deep learning)

Tensor is a generalization of vectors and matrices. Tensor are containers for mostly numerical data. Different tensors are named based on their dimension also known as axis. Dimension are basically the number of indices needed to access a particular entry in a tensor.

1) Scalar (0D tensor): this are just pure numbers. example: 3
2) Vector (1D tensor): example a= [1,2,3,4,5,6]. This example has one axis, the axis has 6 dimensions.it is said to be 1 D vector and 6 Dimensional vector. To access an element a [3] = 4
3) Matrices(2D-tensor):

> A= [[1, 2,3,4], [5, 6,7,8],[9,10,11,12]] to access any element two indices are needed A[2][3] = 11

4)3D tensors and higher dimension tensors: a = [[[1 2], [3 4]], [[5 6], [7 8]]],

To access any entry, we will need 3 indices. a [1][0][1] = 6

The general rule of thumb is when 0D tensors are packed a 1D tensor is formed, packed 1D vectors give a 2D tensor (matrix), packed 2D tensor give a 3D tensor, packed 3D tensors give a 4 D tensor and so on.

Every tensor has 3 key attributes

1)Number of axes(Rank): Which is the dimension of a vector

2) Shape: Is the integer tuple that gives the dimensionality of the tensors from top dimension to the lower dimension

3)Data type: Is the type of data stored in the tensor e. g float32, float64, uint8, int32 and soon

Tensors are what neural network manipulate or operate on at different stage of the network. The input data we have in real world can be

Vector data: (2D tensor) in shape (sample, feature)

Time Series (3D tensor) in (sample, feature, timestamp)

Image 4D tensor in (Sample height, width, channel)

Video 5D tensor in (Sample, frame, height, width, channel)

Tensor operations: In a particular layer of a neural network like below

keras.layers.Dense(512, activation='relu') the activities that occurs there are just tensors operation. Mathematically the above layer can be seen as output = relu (dot(W, input) + b)

the dot operation between w and input, addition operation and the result of dot(W, input) and lastly the relu operation.The operation are considered in category

1) Element-wise Operation: An *element-wise* operation is an operation between two tensors that operates on corresponding elements within the respective tensors. Tensor addition and relu are the element-wise
2) Broadcasting: Broadcasting describes how tensors with different shapes are treated during element-wise operations.
3) Tensor dot
4) Tensor reshaping

Because every tensor can be represented as a point in geometrical space, the operation on tensors also have geometrical interpretation.

Gradient-Based Optimization: output = relu(dot(W, input) + b)

W the weight and b the bias are the trainable parameter of the each layer. Initially they are randomly initialized but as the inputs sample (batch) are passed through and a loss function is used to know how far the output is from the target, there is a need to update these parameters. One way to do update is to compute the **gradient** of the loss with regard to the network's parameters (a *backward pass*)) and move the parameters a little in the opposite direction from the **gradient** thereby reducing the loss on the batch a bit