Name: Most. Masnat Marufi
ID: 20-43122-1
Sub: Computer vision and pattern recognition
Sec: [B]

REPORT ON ACTIVATION FUNCTION

What is Activation Function

The activation function determines whether to activate the neuron. This means that the prediction process uses simpler mathematical operations to determine if a neuron's input to the network is significant.

1. Step function:

A unit step activation function is a commonly used function in neural networks. The output takes the value 0 for negative arguments and 1 for positive arguments. Features include:

$$f(x) = \begin{cases} 0 & \text{when } x<0 \\ 1 & \text{when } x >= 0 \end{cases}$$

The range is between (0,1) and the output is binary in nature. These types of activation functions are useful for binary schemes. If you want to classify your input model into one of two groups, you can use a binary compiler with a unit step enable function.
If the value of Y is above a certain value, declare it valid. If it's less than the threshold, say it's not.

Advantage:

- Great for binary classification.

Disadvantages:

- The slope of the step function is zero. This makes the step function less useful during backward propagation, where the gradient of the activation function is sent for error computation to improve and optimize the result.
- Not available for multiclass classification.

2.Sigmoid function

The sigmoid function is a special form of the logistic function and is usually denoted by f(x) or sig(x). It is given by:

$$f(x) = 1/(1+exp(-x))$$

The sigmoid unit of the neural network. If a neuron's activation function is a sigmoid function, this ensures that the output of that unit is always between 0 and 1. Since the sigmoid function is a nonlinear function, the output of this unit will be a nonlinear function of the weighted sum of the inputs.

Advantage:

- It is a smooth function and continuously differentiable.
- it is non-linear. Therefore, the output is also non-linear.
- Easy to understand and easy to use.
- Calculating derivatives is easy.

Disadvantages:

- The vanishing gradient problem. The sigmoid saturates and kills the gradient.
- The output isn’t zero centered thus the gradient updates go too far in different directions i.e., 0 < output < 1, and it makes optimization harder.
- Sigmoid have slow convergence.

3.Tanh function:

The main advantage of the tanh function is that it aids the backpropagation process by producing a zero-centered output. The tanh function is primarily used in recurrent neural networks for natural language processing and speech recognition tasks. The mathematical formula for the tanh function is follows:

$$f(x) = (e^x – e^{(-x)}) / (e^x + e^{(-x)})$$

Most tanh functions are commonly used in the hidden layers of neural networks. This is because their values range from -1 to 1. This is why the hidden layer means 0 or very close to 0. We center the data by bringing the mean closer to 0, allowing for further training.

Advantages:

- It is continuous and differentiable at all points.
- It basically solves our problem of values all being of the same sign.
- The function as you can see is non-linear so we can easily backpropagate the errors.

Disadvantages:

- Vanishing gradient problem.
- The gradients are low.
- Computationally expensive function

4.ReLU function:

A Rectified Linear Activation Function (ReLU for short) is a piecewise linear function that outputs an input directly if it is positive and zero otherwise. Models that use it are easy to train and often perform better, making it the default activation function for many types of neural networks. The mathematical formula for the ReLU function is as follows:

$$f(x) = \max(0,x)$$

Advantage:

- The ReLU function is nonlinear. This means that errors can be easily backpropagated and multiple layers of neurons are activated by her ReLU function.
- It has been found to significantly speed up the convergence of stochastic gradient descent compared to the sigmoid and tanh functions.
- Not all neurons are activated at the same time. Since some neurons have zero output, few neurons are activated, the network is sparse and efficient, and easy to compute.

Disadvantages:

- It is not differentiable at zero and ReLU is unlimited.

- Negative inputs have zero slope. This means that activations in this region do not update the weights during backpropagation. This can create dead neurons that are never activated. This can be managed by reducing the learning rate and bias.
- The ReLU output is not zero-centered, which degrades neural network performance. The gradients of the weights during backpropagation are either all positive or all negative. This can introduce undesirable zigzag dynamics into the weight gradient updates. This can be handled with batch norm. Batch norm alleviates this problem somewhat by adding these gradients to the entire batch of data, which may change sign with the last update of the weights.
- Average activation is not zero. From ReLU, the network of subsequent layers has a positive bias because the average activation is greater than zero. Its computational simplicity makes it less computationally intensive than Sigmoid and Tanh, but the positive mean shift of the next layer slows down learning.

Always remember that ReLU functions should only be used in hidden layers.

5.ELU function:

An ELU activation layer performs the identity operation on positive inputs and an exponential nonlinearity on negative inputs. The layer performs the following operation: $f(x) = \{ x, x \geq 0 \; \alpha(\exp(x) - 1), x < 0$. The default value of α is 1. Specify a value of α for the layer by setting the Alpha property.

Advantages:

- Gives smoother convergence for any negative axis value.
- For any positive output, it behaves like a step function and gives a constant output.

Disadvantages:

- For x > 0, it can blow up the activation with the output range of [0, inf].

6. SELU function:

The Scaled Exponential Linear Unit (SELU) activation function is defined as: if x > 0: return scale * x. if x < 0: return scale * alpha * (exp(x) - 1) where alpha and scale are pre-defined constants (alpha=1.67326324 and scale=1.05070098).

Advantages:

- Like ReLU, SELU does not have vanishing gradient problem and hence, is used in deep neural networks.
- SELUs learn faster and better than other activation functions without needing further procession. Moreover, other activation functions combined with batch normalization cannot compete with SELUs.

Disadvantages:

- SELU is a relatively new activation function, so it is not yet used widely in practice. ReLU remains the preferred option.
- More research on architectures such as CNNs and RNNs using SELUs is needed for wide-spread industry use.