A

Activation Fxns:

1. Identity
2. Binary step
3. Logistic (aka sigmoid or soft step)
4. TanH
5. ArcTan
6. Softsign
7. Gaussian

Gradient descent algorithms:

1. Batch gradient descent

Gradient descent optimization algorithms

1. Adadelta
2. RMSprop
3. Adam

Algorithms to optimize parallelized and distributed SGD

1. Hogwild!

**Activation Functions:**

1. Identity function:

Advantages:

a. It has infinite range, so training is more efficient.

b. It is a monotonic function with a monotonic derivative, which has been shown to generalize better in some cases and is guaranteed to be convex in single layer model.

c. It is also continuously differentiable, which enables gradient-based optimization methods.

Disadvantages:

a. It is a linear function so multiple layers with identity activation function equal to a single layer model.

2. Binary step function:

Advantages:

a. Add non-linearity to the model because it is not linear.

b. Finite range.

c. It is monotonic, so the error surface is guaranteed to be convex.

Disadvantages:

a. It is not a smooth function with continuous derivative. In detail, it is not differentiable at 0 and its derivative is 0 at all other values so gradient based methods doesn’t work with it.

3. Logistic function:

Advantages:

a. Add non-linearity to the model to confine the output within a specified range.

b. It is continuously differentiable to enable gradient-based optimization.

c. It is monotonic, so the error surface is guaranteed to be convex.

d. Finite range so gradient-based training method is more stable.

e. Used for categorical data.

Disadvantages:

a. It saturates and kill gradients, which will lead no signal flowing through the neuron and the rest of network.

b. The output is not 0 centered

4. Tanh

Advantages:

a. Tanh neuron is a scaled sigmoid neuron. So it has all the advantages that sigmoid function has.

b. The output of Tanh is 0 centered. Return values between -1 and 1.

Disadvantages:

a. Like sigmoid neuron, it saturates and kill gradients.

5. ArcTan

Advantages:

a. ArcTan is similar to sigmoid and Tanh. It squashes inputs between -π/2 and π/2. And it has all the advantages and disadvantages that sigmoid function has.

b. It has a flatter S shape, so it may have a better discrimination between similar values.

c. Output is 0 centered.

Disadvantages:

a. It saturates and kill gradients.

6. Softsign

Advantages:

a. Similar to Tanh (output between -1 and 1), but it might behave differently in terms of saturation because it is polynomial instead of exponential.

Disadvantages:

7. Gaussian:

Advantages/when to use:

a. When expect the final output is a normal distribution, it is suitable to use gaussian as the activation function.

Disadvantages:

a. It is monotonic. Therefore, the error surface is not guaranteed to be convex.

b. The output is not 0 centered. Return values between (0,1].

c. Slower to calculate.

**Gradient descent algorithms:**

Batch gradient descent

Advantages/when to use:

a. When the dataset is small and we can use the whole dataset to compute the gradient of the cost function.

b. It is guaranteed to converge to the global minimum for convex error surface and to a local minimum for non-convex surfaces.

Disadvantages:

a. It is very slow when the dataset is big.

b. doesn’t allow to update model online.

**Gradient descent optimization algorithms**

Adadelta

1. RMSprop
2. Adam

Reference:

1. <https://stats.stackexchange.com/questions/115258/comprehensive-list-of-activation-functions-in-neural-networks-with-pros-cons>

2. <http://cs231n.github.io/neural-networks-1/#classifier>

3. <https://datascience.stackexchange.com/questions/14349/difference-of-activation-functions-in-neural-networks-in-general>

4. <https://en.wikipedia.org/wiki/Hyperbolic_function>

5. <https://visualstudiomagazine.com/articles/2015/06/01/alternative-activation-functions.aspx>

6. <http://proceedings.mlr.press/v9/glorot10a/glorot10a.pdf>

7. <https://stats.stackexchange.com/questions/200888/why-arent-gaussian-activation-functions-used-more-often-in-neural-networks>