

Problem Set 3

Sept. 30, 2021

1. For the multi-layer neural network model

$$g_k(x) \equiv y_k = f\left(\sum_j w_{ij} f\left(\sum_i w_{ij} x_i + w_{j0}\right) + w_{k0}\right),$$

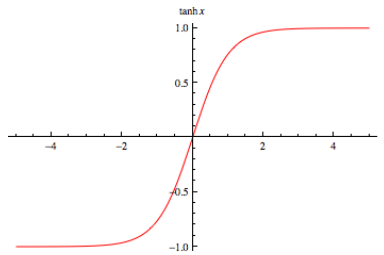
show that this model cannot achieve nonlinearity if the activation function of the neurons $f(\cdot)$ is chosen to be a linear function.

2. For a three-layered MLP model (with one hidden layer)

- (1) use the hyperbolic tangent function

$$\tanh(z) = \frac{\sinh(z)}{\cosh(z)} = \frac{e^z - e^{-z}}{e^z + e^{-z}} = \frac{e^{2z} - 1}{e^{2z} + 1}$$

to replace the sigmoid function as the activation function, derive the BP algorithm for this model and discuss why $\tanh(\cdot)$ is not often used as the activation function.



- (2) The MLP model if for multiple classes and the hidden nodes all use sigmoid activation function, derive the weight-updating formula of the BP algorithm if we replace the sigmoid activation function of output nodes with the SoftMax function:

$$y_j = \frac{e^{\mathbf{w}_j \cdot \mathbf{o}_h}}{\sum_{k=1}^K e^{\mathbf{w}_k \cdot \mathbf{o}_h}}, j = 1, \dots, K$$

where K is the number of classes (therefore the number of output nodes), \mathbf{o}_h is the output vector of the last hidden layer that emits to the output layer, and \mathbf{w}_j is the weight vector of output node j .

Due date: Oct. 6 (Wednesday) 23:00 Beijing time