Basic Deep Learning

1. Deep feedforward networks
2. Hidden layers, without output
3. Loss function/cost function  
   Minimize the loss function to find the best parameters: w, W, c, b
4. Activation function
5. Cost function
6. Example: XOR function

F(x) = w^T max(0, XW + c) + b: define the activation function， because of the activation function.

1. Learning condition distribution with Maximum Likelihood