50.039 Theory and Practice of Deep Learning
Theory Homework 4

Joel Huang 1002530

February 26, 2019

1 Sigmoid function

Derivative of the sigmoid function

$$\varphi(x) = \frac{1}{1+e^{-ax}}$$

$$\varphi(x) = (1+e^{-ax})^{-1}$$

$$\frac{d\varphi}{dx} = (-1)(1+e^{-ax})^{-2}(-ae^{-ax})$$

$$\frac{d\varphi}{dx} = \frac{ae^{-ax}}{(1+e^{-ax})^2}$$

$$\frac{d\varphi}{dx} = a\varphi(x) \cdot \frac{1+e^{-ax}-1}{1+e^{-ax}}$$

$$\frac{d\varphi}{dx} = a\varphi(x) \cdot (1 - \frac{1}{1+e^{-ax}})$$

$$\frac{d\varphi}{dx} = a\varphi(x) \cdot (1-\varphi(x))$$

Value of derivative at $x = 0$

$$\varphi(0) = \frac{1}{2}$$

$$\varphi'(0) = a\varphi(0) \cdot (1-\varphi(0))$$

$$\varphi'(0) = \frac{a}{4}$$

2 Neural network layers

Output of a neural network

The output of the neural network is fed by three inputs, h_1, h_2, w_7.

$$output = g(w_8 h_1 + w_9 h_2 + w_7)$$

The hidden layers are fed by two inputs, x_1, x_2 and a bias each, scaled by the constant c.

$$h_1 = c \cdot (w_3 x_1 + w_5 x_2 + w_1)$$
$$h_2 = c \cdot (w_4 x_1 + w_6 x_2 + w_2)$$

Rearranging,

$$output = g\left(w_8 c \cdot (w_3 x_1 + w_5 x_2 + w_1) + w_9 c \cdot (w_4 x_1 + w_6 x_2 + w_2) + w_7\right)$$

Expressing it in sigmoid form,

$$output = \frac{1}{1+e^{-(w_8 c(w_3 x_1 + w_5 x_2 + w_1) + w_9 c(w_4 x_1 + w_6 x_2 + w_2) + w_7)}}$$

Equation for a binary decision boundary at $g = 0.5$

$$\frac{1}{1+e^{-(w_8 c(w_3 x_1 + w_5 x_2 + w_1) + w_9 c(w_4 x_1 + w_6 x_2 + w_2) + w_7)}} = 0.5$$

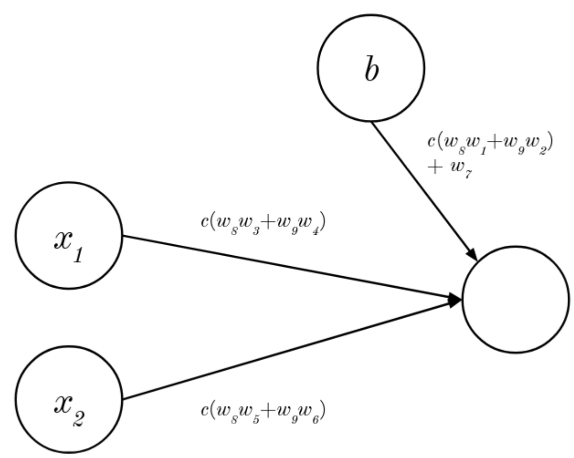
$$e^{-(w_8 c(w_3 x_1 + w_5 x_2 + w_1) + w_9 c(w_4 x_1 + w_6 x_2 + w_2) + w_7)} - 1 = 0$$

$$w_8 c(w_3 x_1 + w_5 x_2 + w_1) + w_9 c(w_4 x_1 + w_6 x_2 + w_2) + w_7 = 0$$

As a function of x_1, x_2:

$$c(w_8 w_3 + w_9 w_4)x_1 + c(w_8 w_5 + w_9 w_6)x_2 + c(w_8 w_1 + w_9 w_2) + w_7 = 0$$

Equivalent neural network without hidden layers, and associated weights



3 Feature map sizes

Dimension of the output is given as:

$$dim_{out} = \left\lceil \frac{dim_{in} - k + 1 + 2p}{s} \right\rceil$$

1. Input
   $(300, 300, 3)$

2. Layer 1
   7×7 conv; 30. $s = 2, p = 0$

   $$dim_{out} = \left\lceil \frac{300 - 7 + 1 + 2(0)}{2} \right\rceil = 147$$

   Layer 1 feature map size: $(147, 147, 30)$

3. Layer 2
   3×3 maxpool. $s = 2, p = 0$

   $$dim_{out} = \left\lceil \frac{147 - 3 + 1 + 2(0)}{2} \right\rceil = 73$$

   Layer 2 feature map size: $(73, 73, 30)$

4. Layer 3
   3×3 conv; 50. $s = 1, p = 3$

   $$dim_{out} = \left\lceil \frac{73 - 3 + 1 + 2(3)}{1} \right\rceil = 77$$

   Layer 3 feature map size: $(77, 77, 50)$