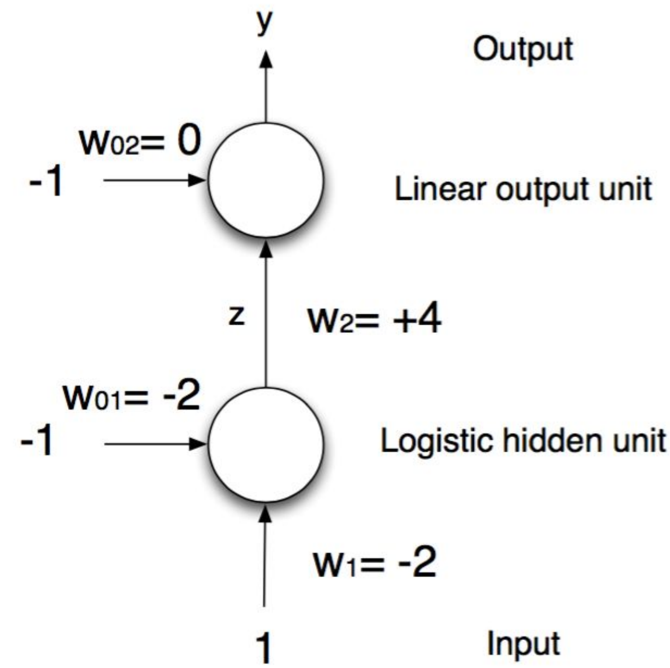
PROBLEM 23

Here you see a very small neural network: it has one input unit, one hidden unit (logistic), and one output unit (linear).



Let's consider one training case. For that training case, the input value is 1 (as shown in the diagram), and the target output value $t = 1$. We're using the following loss function:

$$E = \frac{1}{2}(t - y)^2$$

Please supply numeric answers; the numbers in this question have been constructed in such a way that you don't need a calculator. Show your work in case of mis-calculation in earlier steps.

(a) What is the output of the hidden unit for this input?

Solution: 1/2

(b) What is the output of the output unit for this input?

Solution: 2

(c) What is the loss, for this training case?

Solution: 1/2

(d) What is the derivative of the loss with respect to w_2, for this training case?

Solution: Let z be the output of the hidden unit

$$\begin{aligned}\frac{\partial E}{\partial w_2} &= (1-y)\frac{\partial(-y)}{\partial w_2}\\ &= (1-2)\cdot -z\\ &= (1-2)\cdot -(1/2)\\ &= (1/2)\end{aligned}$$

(e) What is the derivative of the loss with respect to w_1, for this training case?

Solution:

$$\begin{aligned}\frac{\partial E}{\partial w_1} &= \frac{\partial E}{\partial z}\frac{\partial z}{\partial w_1}\\ &= (t-y)\frac{\partial(-y)}{\partial z}\cdot z\cdot(1-z)\cdot x\\ &= (t-y)\cdot -w_2\cdot z\cdot(1-z)\cdot x\\ &= (1-2)\cdot -4\cdot(1/2)\cdot(1/2)\cdot 1\\ &= 1\end{aligned}$$

(f) With sigmoidal activation, the derivative with respect to w_1 and w_2 are

$$\frac{\partial E}{\partial w_2} = -(t-y)z, \text{ and } \frac{\partial E}{\partial w_1} = -(t-y)\cdot w_2\cdot z\cdot(1-z)\cdot x.$$

Assume that we now use the rectified linear unit (ReLU) as our activation (or a *ramp* function). This means that $z = max(0, w_1 x + w_{01})$. What is the derivative of the loss with respect to w_1 and w_2 *at differentiable points* with ReLU? Don't use numerical value for this question.

Solution: It is the same for w_2 as sigmoidal activation case:

$$\frac{\partial E}{\partial w_2} = -(t-y)z$$

For w_1,

$$\begin{aligned}\frac{\partial E}{\partial w_1} &= \frac{\partial E}{\partial z}\frac{\partial z}{\partial w_1}\\ &= (t-y)\frac{\partial(-y)}{\partial z}\cdot z'\cdot x\\ &= -(t-y)\cdot w_2\cdot z'\cdot x,\end{aligned}$$

where,

$$z' = \begin{cases} 1 & \text{if } z > 0\\ 0 & \text{if } z < 0.\end{cases}$$

At $z = 0$, it is not differentiable. For such points, we need to consider subdifferential (or subgradient), but it is not required in this question.

Additional Explanations

a) (−2∗1)+(−2∗−1)=0, and sigmoid(0)=0.5.

b) (0.5∗4)+(0∗−1)=2.

c) E = 0.5∗(2−1)2 = 0.5.