1、 As the function

$$\boldsymbol{y} = \boldsymbol{W}\boldsymbol{x} + \boldsymbol{b}$$

is linear, the property easily follows because the composition of two linear functions is still linear.

2、(1)Acording to the gradient descent method, each time the update will be the form:

$$w_i = w_i - \lambda \frac{\partial E}{\partial w_i}, \quad \forall i$$

while by the chain rule, we have

$$\begin{aligned}
\frac{\partial E}{\partial w_i} &= (y-g)\frac{\partial y}{\partial w_i} \\
&= (y-g)s'(W^T x)x_i \\
&= (y-g)s(W^T x)(1 - s(W^T x))x_i
\end{aligned}$$

So subtituting the second equation to the first and we are done.

(2)We just need to use the formula in (1) for only one update. Here is the answer with the help of calculator:

$$y = s(W^T x) = s(3)$$

$$\boldsymbol{W} = (0.5002, 1.0004, 1.0001)^T$$