1、 If $f(x) = kx + b$， then

$$
\begin{aligned}
g_k(x) \equiv y_k &= f(\sum_j w_{kj} f(\sum_i w_{ji} x_i + w_{j0}) + w_{k0}) \\
&= k_2(\sum_j w_{kj}(k_1(\sum_i w_{ji} x_i + w_{j0}) + b_1) + w_{k0}) + b_2 \\
&= k_2 k_1 \sum_j \sum_i w_{kj} w_{ji} x_i + k_2(\sum_j w_{kj}(k_1 x_i + b_1) + w_{k0}) + b_2 \\
&= w_k x_i + b_k
\end{aligned}
$$

Thus， nonlinearity is not be achieved.

2、(1) assume： $Z = W^T X$

If it use Gradient Descent update weight：

$$
\begin{aligned}
\frac{\partial E}{\partial W} &= \frac{\partial E}{\partial y} \frac{\partial y}{\partial Z} \frac{\partial Z}{\partial W} \\
&= (y - g) y (1 - y) X
\end{aligned}
$$

Then $W = W - \lambda (y - g) y (1 - y) X$

(2) We know W, X, g, λ,

and compute $y = s(W^T X)$,

Based on formula (1), $W = (0.5002, 1.0004, 1.0001)^T$