Homework_1 Part_1：

吕鑫鹏

1：

If the activation function $f(\ )$ is selected as a linear function.

$\sum_{\mathrm{j}} \mathrm{W}_{\mathrm{ji}} \mathrm{X}_{\mathrm{i}} + W_{j0}$ is a linear transformation of X.

So, $f\left(\sum_{\mathrm{j}} \mathrm{W}_{\mathrm{ji}} \mathrm{X}_{\mathrm{i}} + W_{j0}\right)$ also is a linear transformation of X.

In the same way, the next two transformations are also linear, so how many linear transformations to a matrix is equivalent to doing one linear transformation.

Therefore, g(x) is a linear function and that nonlinearity will not be achieved.

2

（1）assume：

$$Z = W^T \bullet X$$

Then,

$$\frac{\partial E}{\partial W^T} = \frac{\partial E}{\partial \mathrm{y}} \bullet \frac{\partial y}{\partial Z} \bullet \frac{\partial Z}{\partial W^T}$$

The update of W can be：

$$W = (W^T - \frac{\partial E}{\partial W^T} \bullet \lambda)^T$$

Using The chained rule:

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

（2）Substituting the given value into the above formula，and e^{-3} values 0.05,

$W = (0.5002, 1.0004, 1.0001)^T$