Assignment 22.

 # Proving that MLPs having all neurons with linear transfer function are just simple 2 layer perzeptrons.

Assume the MLP is 3-layer MLP (N-H-M), we can derive output Y

$$Out_{H_h}=f(net_{H_h})=net_{H_h}=\sum_{n=1}^{N}W_{nh}\cdot X_n=W_{1h}\cdot X_1+W_{2h}\cdot X_2+\cdots+W_{Nh}\cdot X_N$$
$$Out_{Y_m}=f(net_{Y_m})=net_{Y_m}=\sum_{h=1}^{H}U_{hm}\cdot H_h=U_{1m}\cdot H_1+U_{2m}\cdot H_2+\cdots+U_{Hm}\cdot H_H$$
$$Out_{Y_m}=U_{1m}\cdot(W_{11}\cdot X_1+W_{21}\cdot X_2+\cdots+W_{N1}\cdot X_N)+U_{2m}\cdot(W_{12}\cdot X_1+W_{22}\cdot X_2+\cdots+W_{N2}\cdot X_N)+\cdots+U_{Hm}\cdot(W_{1H}\cdot X_1+W_{2H}\cdot X_2+\cdots+W_{NH}\cdot X_N)$$

$$Y_m=(U_{1m}\cdot W_{11}+U_{2m}\cdot W_{12}+\cdots+U_{Hm}\cdot W_{1H})\cdot X_1+(U_{1m}\cdot W_{21}+U_{2m}\cdot W_{22}+\cdots+U_{Hm}\cdot W_{2H})\cdot X_2+\cdots+(U_{1m}\cdot W_{N1}+U_{2m}\cdot W_{N2}+\cdots+U_{Hm}\cdot W_{NH})\cdot X_N$$

Output Y always can be simplified in terms of Input X, where *An* includes the sum of Weights. (regardless of the structure of the network)

$$Y=A_1\cdot X_1+A_2\cdot X_2+\cdots+A_N\cdot X_N$$

This formula represents output of simple 2 layer perzeptron where A1~AN indicates weights to X to Y. What was done during the training process is factorizing *An* into various factors.

 # If linear transfer function will change into $f_2(z)=a*z$, what will change?

Assume that the MLP is consist of 3-layers, then we get

$$Y_m=a^2\cdot(U_{1m}\cdot W_{11}+U_{2m}\cdot W_{12}+\cdots+U_{Hm}\cdot W_{1H})\cdot X_1+a^2\cdot(U_{1m}\cdot W_{21}+U_{2m}\cdot W_{22}+\cdots+U_{Hm}\cdot W_{2H})\cdot X_2$$
$$+\cdots+a^2\cdot(U_{1m}\cdot W_{N1}+U_{2m}\cdot W_{N2}+\cdots+U_{Hm}\cdot W_{NH})\cdot X_N$$

The weights will increase for each additional layer,

$$Y_m=a^{n-1}\cdot(U_{1m}\cdot W_{11}+U_{2m}\cdot W_{12}+\cdots+U_{Hm}\cdot W_{1H})\cdot X_1+a^{n-1}\cdot(U_{1m}\cdot W_{21}+U_{2m}\cdot W_{22}+\cdots+U_{Hm}\cdot W_{2H})\cdot X_2$$
$$+\cdots+a^{n-1}\cdot(U_{1m}\cdot W_{N1}+U_{2m}\cdot W_{N2}+\cdots+U_{Hm}\cdot W_{NH})\cdot X_N$$

where n is number of layer.
So the changing of transfer function by multiplying *a* will change the weights rate for the alternative simple layer perzeptron.

1 Assignment 23

$$
\begin{aligned}
{}^{p}\Delta w_s &= -\eta \frac{\partial^p E^{**}(W_s)}{\partial w_s} \\
{}^{p}E^{**} &= \frac{1}{2}\sum_{m=1}^{M}({}^{p}\hat{Y_m} -^{p} Y_m)^2 + \beta\frac{1}{2}\sum_{i,j}(w_{ij})^2
\end{aligned}
$$

1. Derivation for an output neuron m.

$$
\frac{\partial^p E^{**}}{\partial w_{hm}} = \frac{\partial^p E^{**}}{\partial^p net_m} \cdot \frac{\partial^p net_m}{\partial w_{hm}}
$$

$$
\frac{\partial^p net_m}{\partial w_{hm}} = \frac{\partial}{\partial w_{ij}}\sum_{g=0}^{H} {}^{p}\widetilde{out}_g \cdot w_{gm} =^{p} \widetilde{out}_h
$$

$$
\frac{\partial^p E^{**}}{\partial^p net_m} = \frac{\partial^p E^{**}}{\partial^p y_m} \cdot \frac{\partial^p y_m}{\partial^p net_m}
$$

$$
\frac{\partial^p y_m}{\partial^p net_m} = \frac{\partial f({}^{p}net_m)}{\partial^p net_m} = f^{'}({}^{p}net_m)
$$

$$
\begin{aligned}
\frac{\partial^p E^{**}}{\partial^p y_m} &= \frac{\partial}{\partial^p y_m}\frac{1}{2}\sum_{j=1}^{M}({}^{p}\hat{y_j} -^{p} y_j)^2 + \frac{\partial}{\partial^p y_m}\beta\frac{1}{2}\sum_{i,j}(w_{ij})^2 \\
&= \frac{1}{2}\frac{\partial}{\partial^p y_m}({}^{p}\hat{y_m} -^{p} y_m)^2 + 0 \\
&= \frac{1}{2}2({}^{p}\hat{y_m} -^{p} y_m)\frac{\partial}{\partial^p y_m}(-y_m) \\
&= ({}^{p}\hat{y_m} -^{p} y_m)\cdot(-1) \\
&= -({}^{p}\hat{y_m} -^{p} y_m)
\end{aligned}
$$

Now, we get weight change formula - delta-rule.

$$
\begin{aligned}
{}^{p}\Delta w_s &= \eta \cdot ({}^{p}\hat{y_m} -^{p} y_m) \cdot f^{'}({}^{p}net_m) \cdot^{p} \widetilde{out}_h \\
{}^{p}\Delta w_{hm} &= \eta \cdot^{p} \delta_m^{**} \cdot^{p} \widetilde{out}_h \\
&\quad where,\ \ {}^{p}\delta_m^{**} = -\frac{\partial^p E^{**}}{\partial^p net_m}
\end{aligned}
$$

2. Derivation for a hidden neuron h.

$$
{}^{p}\Delta w_{gh} = \eta \cdot^{p} \delta_h^{**} \cdot^{p} \widetilde{out}_g
$$

$$
\begin{aligned}
-{}^{p}\delta_h^{**} &= \frac{\partial^p E^{**}}{\partial^p net_h} \\
&= \frac{\partial^p E^{**}}{\partial^p out_h} \cdot f^{'}({}^{p}net_h) \\
&= \sum_{k=1}^{K}\left(\frac{\partial^p E^{**}}{\partial \underline{net_k}}\frac{\partial \underline{net_k}}{\partial^p out_h}\right) \cdot f^{'}({}^{p}net_h) \\
&= -\sum_{k=1}^{K}(\underline{\delta^{**}}_k \cdot \underline{w_{hk}}) \cdot f^{'}({}^{p}net_h)
\end{aligned}
$$