Derivation of the Backpropagation Learning for network with one hidden layer.

- **Notation:**
  - I = the number of input units
  - J = the number of hidden units
  - K = the number of output units
  - v_{ji} is the weight of the connection from input unit $i = 1, 2, \ldots I$ to hidden unit $j = 1, 2, \ldots, J$.
  - w_{kj} is the weight of the connection from hidden unit $j = 1, 2, \ldots J$ to output unit $k = 1, 2, \ldots, K$
  - $\mathbf{z}$ is the input vector with components $z_1, z_2, \ldots, z_I$.
  - $\mathbf{d}$ is the output vector with components $d_1, d_2, \ldots, d_K$.
- **Activation at the hidden unit j:**

$$\begin{aligned} y_j &= f(net_j) &(1)\\ net_j &= \sum_{i=1}^{I} v_{ji} z_i &(2) \end{aligned}$$

- **Activation at the output unit k:**

$$\begin{aligned} o_k &= f(net_k) &(3)\\ net_k &= \sum_{j=1}^{J} w_{kj} y_j &(4) \end{aligned}$$

- **Total error at the output units:**

$$E = \frac{1}{2} \sum_{k=1}^{K} (d_k - o_k)^2$$

- **Computation of w_{kj} adjustment:**
  - Weight adjusment:

$$\Delta w_{kj} = -\eta \frac{\partial E}{\partial w_{kj}} \tag{5}$$

  - Take the partial derivative of E with respect to w_{kj}:

$$\begin{aligned} \frac{\partial E}{\partial w_{kj}} &= \frac{\partial \frac{1}{2} \sum_{k=1}^{K} (d_k - o_k)^2}{\partial w_{kj}} &(6)\\ &= \frac{\partial \frac{1}{2} (d_k - o_k)^2}{\partial w_{kj}} &(7)\\ &= -(d_k - o_k) \times \frac{\partial o_k}{\partial w_{kj}} &(8) \end{aligned}$$

$$= -(d_k - o_k) \times \frac{\partial f(net_k)}{\partial w_{kj}} \tag{9}$$

$$= -(d_k - o_k) \times \frac{\partial f(\sum_{j=1}^{J} w_{kj} y_j)}{\partial w_{kj}} \tag{10}$$

$$= -(d_k - o_k) \times f^{'}(\sum_{j=1}^{J} w_{kj} y_j) \times \frac{\partial (\sum_{j=1}^{J} w_{kj} y_j)}{\partial w_{kj}} \tag{11}$$

$$= -(d_k - o_k) \times f^{'}(\sum_{j=1}^{J} w_{kj} y_j) \times y_j \tag{12}$$

- Let

$$\delta_{ok} = (d_k - o_k) \times f^{'}(\sum_{j=1}^{J} w_{kj} y_j) \tag{13}$$

$$= (d_k - o_k) \times f^{'}(net_k) \tag{14}$$

$$= (d_k - o_k) \times o_k^{'} \tag{15}$$

then

$$\frac{\partial E}{\partial w_{kj}} = \frac{\partial \frac{1}{2} \sum_{k=1}^{K} (d_k - o_k)^2}{\partial w_{kj}} \tag{16}$$

$$= -(d_k - o_k) \times f^{'}(\sum_{j=1}^{J} w_{kj} y_j) \times y_j \tag{17}$$

$$= -\delta_{ok} \times y_j \tag{18}$$

and

$$\Delta w_{kj} = -\eta \frac{\partial E}{\partial w_{kj}} \tag{19}$$

$$= \eta \, \delta_{ok} \, y_j \tag{20}$$

- **Computation of v_{ji} adjustment:**
  - Weight adjustment:

$$\Delta v_{ji} = -\eta \frac{\partial E}{\partial v_{ji}} \tag{21}$$

  - Take the partial derivative of E with respect to v_{ji}:

$$\frac{\partial E}{\partial v_{ji}} = \frac{\partial \frac{1}{2} \sum_{k=1}^{K} (d_k - o_k)^2}{\partial v_{ji}} \tag{22}$$

$$= -\sum_{k=1}^{K} (d_k - o_k) \frac{\partial o_k}{\partial v_{ji}} \tag{23}$$

$$= -\sum_{k=1}^{K} (d_k - o_k) \frac{\partial f(net_k)}{\partial v_{ji}} \tag{24}$$

$$
\begin{aligned}
&= -\sum_{k=1}^{K}(d_k-o_k)\, f^{'}(net_k)\frac{\partial(net_k)}{\partial v_{ji}} &(25)\\
&= -\sum_{k=1}^{K}(d_k-o_k)\, f^{'}(net_k)\frac{\partial\sum_{j=1}^{J}w_{kj}y_j}{\partial v_{ji}} &(26)\\
&= -\sum_{k=1}^{K}(d_k-o_k)\, f^{'}(net_k)\frac{\partial w_{kj}y_j}{\partial v_{ji}} &(27)\\
&= -\sum_{k=1}^{K}(d_k-o_k)\, f^{'}(net_k)\frac{\partial w_{kj}f(net_j)}{\partial v_{ji}} &(28)\\
&= -\sum_{k=1}^{K}(d_k-o_k)\, f^{'}(net_k)w_{kj}f^{'}(net_j)\frac{\partial\sum_{i=1}^{I}net_j}{\partial v_{ji}} &(29)\\
&= -\sum_{k=1}^{K}(d_k-o_k)\, f^{'}(net_k)w_{kj}f^{'}(net_j)\frac{\partial\sum_{i=1}^{I}v_{ji}z_i}{\partial v_{ji}} &(30)\\
&= -\sum_{k=1}^{K}(d_k-o_k)\, f^{'}(net_k)w_{kj}f^{'}(net_j)z_i &(31)\\
&= -\sum_{k=1}^{K}\delta_{ok}\, w_{kj}f^{'}(net_j)z_i &(32)
\end{aligned}
$$

- Let

$$
\begin{aligned}
\delta_{yj} &= \sum_{k=1}^{K}\delta_{ok}w_{kj}f^{'}(net_j) &(33)\\
&= \sum_{k=1}^{K}\delta_{ok}\times w_{kj}\times y_j^{'} &(34)
\end{aligned}
$$

then

$$
\begin{aligned}
\frac{\partial E}{\partial v_{ji}} &= \frac{\partial\frac{1}{2}\sum_{k=1}^{K}(d_k-o_k)^2}{\partial v_{ji}} &(35)\\
&= -\sum_{k=1}^{K}\delta_{ok}\, w_{kj}\, f^{'}(net_j)\, z_i &(36)\\
&= -\delta_{yj}\, z_i &(37)
\end{aligned}
$$

and

$$
\begin{aligned}
\Delta v_{ji} &= -\eta\,\frac{\partial E}{\partial v_{ji}} &(38)\\
&= \eta\,\delta_{yj}\, z_i &(39)
\end{aligned}
$$