1. For a network layer with n inputs and a maxout activation ϕ, it is formally

$$\phi(x) = \max\left(\langle w_i^T x + b_i \rangle_{i=1}^{n}\right)$$

For backprop,

$y_n \rightarrow$ ground truth

$d_n = \phi(x)$

$\varepsilon_n = \frac{1}{2} \| y_n - d_n \|^2$

We need to update weight ω_n, by $\Delta\omega_n$

$$\Delta\omega_n = -\eta \frac{\partial \varepsilon_n}{\partial \omega_n} \quad , \eta \text{ being the learning rate}$$

$$= -\eta \frac{\partial}{\partial \omega_n} \| \phi(x_n) - y_n \|^2 \cdot \frac{1}{2}$$

$$= -\eta \, \varepsilon_n \cdot \frac{\partial}{\partial \omega_n} \phi(x_n)$$

$$= \begin{cases} -\eta \varepsilon_n x_n & n = i \quad \text{s.t.} \quad w_i x + b = \phi(x) \\ 0 & n = i \quad \text{s.t.} \quad w_i x + b < \phi(x) \end{cases}$$

Ans.