1.

$$\sigma'(\cdot)=\sigma(\cdot)(1-\sigma(\cdot))$$

(1)

对于回归问题

$$\frac{\partial y_k}{\partial w_{kj}}=z_j$$

$$\frac{\partial y_k}{\partial w_{ji}}=\frac{\partial a_k}{\partial z_j}\frac{\partial z_j}{\partial a_j}\frac{\partial a_j}{\partial w_{ji}}=w_{kj}h'(a_j)z_i$$

对于分类问题

$$\frac{\partial y_k}{\partial w_{kj}}=\frac{\partial y_k}{\partial a_k}\frac{\partial a_k}{\partial w_{kj}}=\sigma'(a_k)w_{kj}$$

$$\frac{\partial y_k}{\partial w_{ji}}=\frac{\partial y_k}{\partial a_k}\frac{\partial a_k}{\partial z_j}\frac{\partial z_j}{\partial a_j}\frac{\partial a_j}{\partial w_{ji}}=\sigma'(a_k)w_{kj}h'(a_j)z_i$$

(2)

对于回归问题

$$\frac{\partial E_n}{\partial w_{kj}}=\frac{\partial E_n}{\partial y_k}\cdot\frac{\partial y_k}{\partial w_{kj}}=\delta_k z_j$$

$$\frac{\partial E_n}{\partial w_{ji}}=\frac{\partial E_n}{\partial y_k}\cdot\frac{\partial y_k}{\partial z_j}\cdot\frac{\partial z_j}{\partial w_{ji}}=\delta_j z_i$$

其中

$$\delta_k=\frac{\partial E_n}{\partial y_k}=y_k-t_k$$

$$\delta_j=h'(a_j)\sum_k w_{kj}\delta_k$$

对于分类问题，形式相同，但是

$$\delta_k=\frac{\partial E_n}{\partial y_k}=-\frac{t_k}{y_k}+\frac{1-t_k}{1-y_k}$$

(3)

对于回归问题

$$\frac{\partial y_k}{\partial z_i}=\frac{\partial a_k}{\partial z_j}\frac{\partial z_j}{\partial a_j}\frac{\partial a_j}{\partial z_i}=w_{kj}h'(a_j)w_{ji}$$

对于分类问题

$$\frac{\partial y_k}{\partial z_i}=\frac{\partial y_k}{\partial a_k}\frac{\partial a_k}{\partial z_j}\frac{\partial z_j}{\partial a_j}\frac{\partial a_j}{\partial z_i}=\sigma'(a_k)w_{kj}h'(a_j)w_{ji}$$

2.

(1)

对于回归

$$
\begin{aligned}
w_{MAP} = (\Sigma_0^{-1} + \frac{1}{\sigma^2}X^TX)^{-1}(\Sigma_0^{-1}m_0 + \frac{1}{\sigma^2}X^T\mathbf{t}) \\
p(w|D) \propto p(D|w)p(w) \\
\propto \prod_{n=1}^{N} p(t_n|x_n, w)\mathcal{N}(m_0, \Sigma_0^{-1})
\end{aligned}
$$

对于分类

使用逻辑回归时，似然函数会是一个关于交叉熵的函数。后验概率$p(w|D)$的最大化通常不能直接解析求解，需要使用数值优化方法。

(2)

对于分类

$$
\begin{aligned}
p(t_{N+1}|x_{N+1}, D) &= p(t_{N+1}|x_{N+1}, \theta_{MAP}) \\
&= y(x_{N+1}, \theta_{MAP})^{t_{N+1}}[1 - y(x_{N+1}, \theta_{MAP})]^{1-t_{N+1}}
\end{aligned}
$$

对于回归

$$
p(t_{N+1}|x_{N+1}, D) \sim \mathcal{N}(y(t_{N+1}, \theta_{MAP}), \overline{g}_{MAP}^T H_{MAP}^{-1} \overline{g}_{MAP} + \beta^{-1})
$$