## 5.2

$\Delta w_i = \eta(y-\hat{y})x_i$ [推导]：此处感知机的模型为： $y=f(\sum_i w_i x_i - \theta)$ 将$\theta$看成哑结点后，模型可化简为：
$y=f(\sum_i w_i x_i)=f(\boldsymbol{w}^T\boldsymbol{x})$ 其中$f$为阶跃函数。
根据《统计学习方法》§2可知，假设误分类点集合为$M$，$\boldsymbol{x}_i\in M$为误分类点，$\boldsymbol{x}_i$的真实标签为$y_i$,模型的预测值为$\hat{y}_i$
,对于误分类点$\boldsymbol{x}_i$来说，此时$\boldsymbol{w}^T\boldsymbol{x}_i>0,\hat{y}_i=1,y_i=0$或$\boldsymbol{w}^T\boldsymbol{x}_i<0,\hat{y}_i=0,y_i=1$,综合考虑两种情形可得：
$(\hat{y}_i-y_i)\boldsymbol{w}\boldsymbol{x}_i>0$ 所以可以推得损失函数为： $L(\boldsymbol{w})=\sum_{\boldsymbol{x}_i\in M}(\hat{y}_i-y_i)\boldsymbol{w}\boldsymbol{x}_i$ 损失函数的梯度为：
$\nabla_w L(\boldsymbol{w})=\sum_{\boldsymbol{x}_i\in M}(\hat{y}_i-y_i)\boldsymbol{x}_i$ 随机选取一个误分类点$(\boldsymbol{x}_i,y_i)$，对$\boldsymbol{w}$进行更新：
$\boldsymbol{w}\leftarrow\boldsymbol{w}-\eta(\hat{y}_i-y_i)\boldsymbol{x}_i=\boldsymbol{w}+\eta(y_i-\hat{y}_i)\boldsymbol{x}_i$ 显然式5.2为$\boldsymbol{w}$的第$i$个分量$w_i$的变化情况

## 5.12

$\Delta\theta_j=-\eta g_j$ [推导]：因为 $\Delta\theta_j=-\eta\dfrac{\partial E_k}{\partial\theta_j}$ 又

$$\begin{aligned}\frac{\partial E_k}{\partial\theta_j}&=\frac{\partial E_k}{\partial\hat{y}_j^k}\cdot\frac{\partial\hat{y}_j^k}{\partial\theta_j}\\&=(\hat{y}_j^k-y_j^k)\cdot f'(\beta_j-\theta_j)\cdot(-1)\\&=-(\hat{y}_j^k-y_j^k)f'(\beta_j-\theta_j)\\&=g_j\end{aligned}$$

所以 $\Delta\theta_j=-\eta\dfrac{\partial E_k}{\partial\theta_j}=-\eta g_j$

## 5.13

$\Delta v_{ih}=\eta e_h x_i$ [推导]：因为 $\Delta v_{ih}=-\eta\dfrac{\partial E_k}{\partial v_{ih}}$ 又

$$\begin{aligned}\frac{\partial E_k}{\partial v_{ih}}&=\sum_{j=1}^{l}\frac{\partial E_k}{\partial\hat{y}_j^k}\cdot\frac{\partial\hat{y}_j^k}{\partial\beta_j}\cdot\frac{\partial\beta_j}{\partial b_h}\cdot\frac{\partial b_h}{\partial\alpha_h}\cdot\frac{\partial\alpha_h}{\partial v_{ih}}\\&=\sum_{j=1}^{l}\frac{\partial E_k}{\partial\hat{y}_j^k}\cdot\frac{\partial\hat{y}_j^k}{\partial\beta_j}\cdot\frac{\partial\beta_j}{\partial b_h}\cdot\frac{\partial b_h}{\partial\alpha_h}\cdot x_i\\&=\sum_{j=1}^{l}\frac{\partial E_k}{\partial\hat{y}_j^k}\cdot\frac{\partial\hat{y}_j^k}{\partial\beta_j}\cdot\frac{\partial\beta_j}{\partial b_h}\cdot f'(\alpha_h-\gamma_h)\cdot x_i\\&=\sum_{j=1}^{l}\frac{\partial E_k}{\partial\hat{y}_j^k}\cdot\frac{\partial\hat{y}_j^k}{\partial\beta_j}\cdot w_{hj}\cdot f'(\alpha_h-\gamma_h)\cdot x_i\\&=\sum_{j=1}^{l}(-g_j)\cdot w_{hj}\cdot f'(\alpha_h-\gamma_h)\cdot x_i\\&=-f'(\alpha_h-\gamma_h)\cdot\sum_{j=1}^{l}g_j\cdot w_{hj}\cdot x_i\\&=-b_h(1-b_h)\cdot\sum_{j=1}^{l}g_j\cdot w_{hj}\cdot x_i\\&=-e_h\cdot x_i\end{aligned}$$

所以 $\Delta v_{ih} = -\eta \cdot -e_h \cdot x_i = \eta e_h x_i$

# 5.14

$\Delta \gamma_h = -\eta e_h$ [推导]：因为 $\Delta \gamma_h = -\eta \frac{\partial E_k}{\partial \gamma_h}$ 又

$$\begin{aligned}
\frac{\partial E_k}{\partial \gamma_h} &= \sum_{j=1}^{l} \frac{\partial E_k}{\partial \hat{y}_j^k} \cdot \frac{\partial \hat{y}_j^k}{\partial \beta_j} \cdot \frac{\partial \beta_j}{\partial b_h} \cdot \frac{\partial b_h}{\partial \gamma_h} \\
&= \sum_{j=1}^{l} \frac{\partial E_k}{\partial \hat{y}_j^k} \cdot \frac{\partial \hat{y}_j^k}{\partial \beta_j} \cdot \frac{\partial \beta_j}{\partial b_h} \cdot f'(\alpha_h - \gamma_h) \cdot (-1) \\
&= -\sum_{j=1}^{l} \frac{\partial E_k}{\partial \hat{y}_j^k} \cdot \frac{\partial \hat{y}_j^k}{\partial \beta_j} \cdot w_{hj} \cdot f'(\alpha_h - \gamma_h) \\
&= e_h
\end{aligned}$$

所以 $\Delta \gamma_h = -\eta e_h$