

图3.5给出了感知器参数学习的更新过程，其中红色实心点为正例，蓝色空心点为负例。黑色箭头表示当前的权重向量，红色虚线箭头表示权重的更新方向。

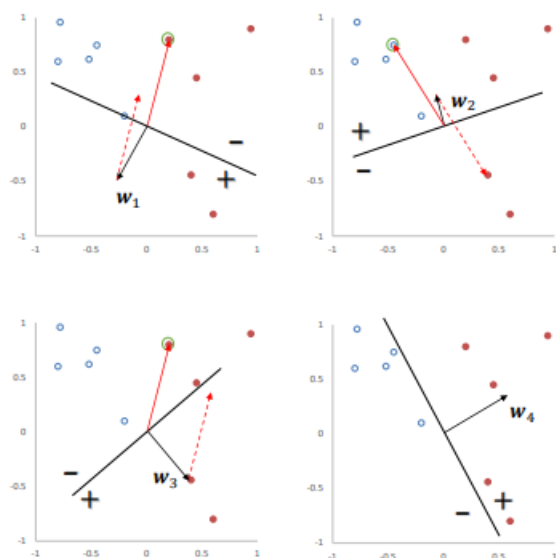


图 3.5 感知器参数学习的更新过程

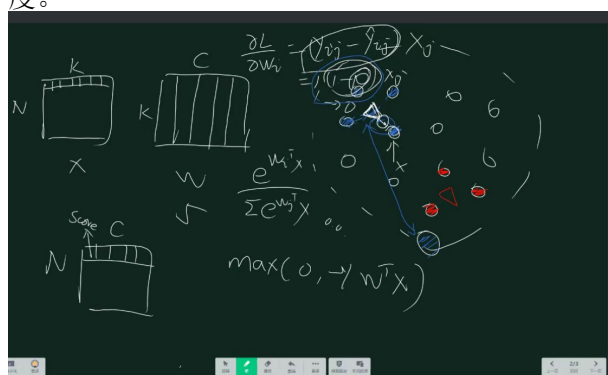
根据感知器的学习策略,可以反推出感知器的损失函数为

$$\mathcal{L}(w; x, y) = \max(0, -yw^T x) \quad (3.57)$$

采用随机梯度下降,其每次更新的梯度为

$$\frac{\partial \mathcal{L}(w; x, y)}{\partial w} = \begin{cases} 0 & \text{if } yw^T x > 0, \\ -yx & \text{if } yw^T x < 0. \end{cases} \quad (3.58)$$

他前面说 Softmax 回归中 w 的更新过程与《神经网络与深度学习》第三章里面的感知器类似。他画了一个类似图 3.5 的图,说在感知器参数学习的更新过程中,只有被错分的样本会使参数有一个更新的梯度。



在这张图里面,每一个圆圈表示一个样本,每一个三角形表示一个类别,预测每个样本的类别就是离它最近的三角形。表示类别的三角形位置移动就表示 w 的更新。

根据他写的梯度的这个公式,对于 p 比较大的样本,它产生的梯度就较小;对于 p 比较小的样本(预测的比较不准),它产生的梯度就较大,就会使这个三角形倾向于向这个样本的方向移动。