3，总结梯度下降法、随机梯度下降法、Adagrad、RMSProp、动量法（Momentum）和 Adam 等方法权系数更新表达式。

解：对于任意的损失函数 L，假设任一单个样本 n 的梯度$\nabla L_n(\boldsymbol{w})$，$t$ 代表迭代次数

（1）梯度下降法：

$$\nabla L_{in}(\boldsymbol{w}) = \frac{1}{N}\sum_{n=1}^{N} \nabla L_n(\boldsymbol{w})$$

$$\boldsymbol{w}_{t+1} \leftarrow \boldsymbol{w}_t - \eta \nabla L_{in}(\boldsymbol{w}_t)$$

（2）随机梯度下降法：

$\nabla L_{in}(\boldsymbol{w}) = \frac{1}{B}\sum_{n=1}^{B} \nabla L_n(\boldsymbol{w})$，$B$ 代表批量大小，最小可以为 1

$$\boldsymbol{w}_{t+1} \leftarrow \boldsymbol{w}_t - \eta \nabla L_{in}(\boldsymbol{w}_t)$$

（3）Adagrad：

$$\nabla L_{in}(\boldsymbol{w}) = \frac{1}{B}\sum_{n=1}^{B} \nabla L_n(\boldsymbol{w})$$

$\sigma_t = \sqrt{\frac{1}{t+1}\sum_{t=0}^{t}(\nabla L_{in}(\boldsymbol{w}))^2 + \varepsilon}$， ε代表极小量，防止σ_t为 0

$$\boldsymbol{w}_{t+1} \leftarrow \boldsymbol{w}_t - \frac{\eta}{\sigma_t} \nabla L_{in}(\boldsymbol{w}_t)$$

（4）RMSProp：

$$\nabla L_{in}(\boldsymbol{w}) = \frac{1}{B}\sum_{n=1}^{B} \nabla L_n(\boldsymbol{w})$$

$$\sigma_{t-1} = \sqrt{\frac{1}{t}\sum_{t=0}^{t-1}(\nabla L_{in}(\boldsymbol{w}))^2}$$

$$\sigma_t = \sqrt{\alpha(\sigma_{t-1})^2 + (1-\alpha)\left(\nabla L_{in}(\boldsymbol{w})\right)^2 + \varepsilon}$$

$$\boldsymbol{w}_{t+1} \leftarrow \boldsymbol{w}_t - \frac{\eta}{\sigma_t} \nabla L_{in}(\boldsymbol{w}_t)$$

（5）动量法（Momentum）：