1. 对于函数 $f(x,y)=25x^2+y^2$, 计算:

a. 在初始点 $c=(0.6,4)$ 处的梯度向量。

b. 在初始点 $(0.6,4)$ 处的标准化向量。

c. 如果取步长为 0.5 , 那么按照梯度下降法, 初始点随向量 c 的方向下降, 计算在下一步迭代的点的位置。

d. 如果取步长为 0.5 , 梯度下降方向是 $(1,0)^T$, 计算下一步迭代点的位置。

首先计算函数 $f(x,y)=25x^2+y^2$ 的梯度向量:

$$\nabla f(x,y)=(\partial f/\partial x,\partial f/\partial y)=(50x,2y)$$

a. 在初始点 c $=(0.6,4)$ 处的梯度向量:

$$\nabla f(0.6,4)=(50*0.6,2*4)=(30,8)$$

b. 在初始点 $(0.6,4)$ 处的标准化向量: 首先计算梯度向量的长度:

$$\|\nabla f(0.6,4)\|=\sqrt{(30^2+8^2)}=\sqrt{(900+64)}=\sqrt{964}\approx 31.0483$$

然后，求梯度向量的单位向量：标准化向量 $=(30/31.0483,8/31.0483)\approx(0.9662,0.2576)$

c. 新的点 = 初始点 - 步长 * 梯度向量新的点 $=(0.6,4)-0.5*(30,8)=(0.6-15,4-4)=(-14.4,0)$ 所以，下一步迭代的点的位置为 $(-14.4,0)$ 。

d. 如果取步长为 0.5 ，梯度下降方向是 $(1,0)^T$ ，计算下一步迭代点的位置: 新的点 = 初始点 - 步长 * 梯度下降方向新的点 $=(0.6,4)-0.5*(1,0)=(0.6-0.5,4)=(0.1,4)$ 所以，下一步迭代点的位置为 $(0.1,4)$ 。

4. 根据随机梯度下降法的定义, 写出最小化问题

$$\min_{P^*,Q^*} J(R;P,Q)=\frac{1}{2}\sum_{(u,i)\in K}\left(r_{ui}-p_u^T q_i\right)^2$$

中参数 P, Q 的更新公式。

根据定义，需要对损失函数 $J(R; P, Q)$ 求关于 P 和 Q 的梯度，并使用这些梯度进行参数更新。首先将损失函数分为两部分：预测误差平方和与正则化项。为了简化表示，我们用 E 表示误差项 $(r_u i - p_u^T q_i)$ ，则损失函数 $J(R; P, Q)$ 可以表示为:

$$J(R; P, Q) = 1/2 * \Sigma_{(u,i)\in K} \left(E^2 + \lambda \left(\|P\|_2^2 + \|Q\|_2^2\right)\right)$$

计算损失函数关于 P 和 Q 的偏导数。

$\partial J/\partial p_u = -E * q_i + \lambda * p_u$

$\partial J/\partial q_i = -E * p_u + \lambda * q_i$

使用学习率 (记为 ϵ) 来更新参数 P 和 Q:

$$p_u' = p_u - \epsilon * \partial J/\partial p_u = p_u + \epsilon * (E * q_i - \lambda * p_u)$$
$$q_i' = q_i - \epsilon * \partial J/\partial q_i = q_i + \epsilon * (E * p_u - \lambda^* q_i)$$

其中， p_u' 和 q_i' 分别表示更新后的参数。

5. 设矩阵分解的重构误差函数为

$$J(R; P, Q) = \frac{1}{2} \sum_{(u,i)\in K} \left(r_{ui} - p_u^T q_i\right)^2$$

其中, r_{ui} 是用户 u 对项目 i 的真实评分, $\widehat{r}_{ui} = p_u^T q_i$ 是模型的估计值, K 为评分矩阵 R 中能够观测到评分的用户-项目对, 即 $K = \{(u, i) \mid r_{ui} \neq 0\}$ 。

a. 证明: $J(R; P, Q)$ 不是关于 P 和 Q 的凸函数;

b. 计算 $\frac{\partial J}{\partial p_{uj}}$ 和 $\frac{\partial J}{\partial q_{ji}}$ 。

c. 当我们使用随机梯度下降算法学习参数时, 写出参数 p_{uj} 和参数 q_{ji} 的更新规则。

a. 可以证明 Hessian 矩阵不是半正定阵或者直接举反例。

b.

$$\partial J/\partial p_{uj} = -\Sigma_{(u,j)\in K} \left(r_{ui} - p_u^T q_i\right) * q_{ij}$$

$$\partial J/\partial q_{ji} = -\Sigma_{(u,j)\in K}\left(r_{ui} - p_u^T q_i\right) * p_{uj}$$

c. 学习率 ϵ 进行参数更新。

$$p'_{uj} = p_{uj} + \epsilon * \left(r_{ui} - p_u^T q_i\right) * q_{ij}$$
$$p'_{uj} = q^{ji} + \epsilon * \left(r_{ui} - p_u^T q_i\right) * p_{uj}.$$

其中 p'_{uj} 和 p'_{uj} 分别表示更新后的参数。