《大数据机器学习》第 5 次作业

姓名：刘培源　　学号：2023214278

题目 1： 写出逻辑斯谛回归模型学习的梯度下降算法。

答：

设逻辑回归的假设为：

$$h_\theta(x) = P(Y=1|x) = \frac{1}{1+e^{-\theta x}}$$

其对应的损失函数为：

$$L(\theta) = \sum_i (-y^{(i)} \log(h_\theta(x^{(i)})) - (1-y^{(i)}) \log(1-h_\theta(x^{(i)})))$$

对损失函数求梯度，我们得到：

$$g(\theta) = \nabla L(\theta) = \sum_i (h_\theta(x^{(i)}) - y^{(i)}) x^{(i)}$$

在梯度下降法中，我们要找到最小化损失函数的参数 。为此，我们需要迭代地更新参数：

1. 初始化参数：$\theta^{(0)}$, 设定迭代次数为 k=0
2. 计算梯度：$g(\theta^{(k)})$
3. 更新参数: 如果 $\|g(\theta^{(k)})\| < \epsilon$, 则令 $\theta^* = \theta^{(k)}$ 并停止; 否则, $\theta^{(k+1)} = \theta^{(k)} - \eta g(\theta^{(k)})$
4. 检查收敛条件：如果 $\|\theta^{(k+1)} - \theta^{(k)}\| < \epsilon'$，则令 $\theta^* = \theta^{(k)}$ 并停止
5. 如果未满足上述任何条件，将 k 更新为 k+1，并返回第 2 步

其中，θ 是模型参数，$h_\theta(x)$ 是假设函数，$L(\theta)$ 是损失函数，$g(\theta)$ 是损失函数的梯度，η 是学习率，ϵ 和 ϵ' 是预定义的两个阈值，用于检查梯度和参数更新的收敛条件。

题目 2：写出最大熵模型学习的 DFP 算法。

答：

最大熵模型的概率分布为：

$$P_u(y|x) = \frac{1}{Z_u(x)} \exp\left(\sum_{i=1}^{n} u_i f_i(x,y)\right)$$

其中规范化因子是

$$Z_u(x) = \sum_y \exp\left(\sum_{i=1}^{n} u_i f_i(x,y)\right)$$

最大熵模型的期望值与经验值之间的差为：

$$f(u) = \sum_y P(x,y) \log \frac{\sum_y \exp\left(\sum_{i=1}^{n} u_i f_i(x,y)\right)}{\sum_{x,y} P(x,y) \sum_{i=1}^{n} u_i f_i(x,y)}$$

其中的梯度为：

$$g(u) = \left(\frac{\partial f(u)}{\partial u_1}, \frac{\partial f(u)}{\partial u_2}, ..., \frac{\partial f(u)}{\partial u_n}\right)^T$$

$$g(u) = \sum_{x,y} P(x) P_u(y|x) f(x,y) - E_P(f_i)$$

在 DFP 算法中，我们用四个向量 H_k, g_k, p_k, 和 q_k 表示。DFP 算法主要用于寻找一个向量 u 使得 $f(u)$ 取得最小值。具体的迭代公式如下：

$$u_k = g_k^{-1} g_k$$

$$\delta_k = u_{k+1} - u_k$$

$$\gamma_k = g_{k+1} - g_k$$

在迭代过程中，我们有：

$$H_{k+1} = H_k + \frac{\delta_k \delta_k^T}{\delta_k^T \gamma_k} - \frac{H_k \gamma_k \gamma_k^T H_k}{\gamma_k^T H_k \gamma_k}$$

最大熵模型学习的 DFP 算法如下：

1. 初始化参数 u_0，并设定一个小的正值 ϵ。
2. 计算梯度 $g_0 = g(u_0)$。
3. 如果 $||g_k|| < \epsilon$，则结束。
4. 选择一个搜索方向 $p_k = -H_k g_k$。
5. 使用线搜索找到一个合适的步长 α_k，使得 $f(u_k + \alpha_k p_k)$ 取得最小值。
6. 更新参数 $u_{k+1} = u_k + \alpha_k p_k$。
7. 更新梯度 $g_{k+1} = g(u_{k+1})$ 和矩阵 H_{k+1}。
8. 若满足终止条件则结束，否则返回第 3 步。