1. 请写出logistic regression的p_i的解析式

logistic regression的p_i是指事件发生的概率，它可以用以下公式表示：

$$p_i = \frac{1}{1+e^{-z_i}}$$

其中z_i是logit变换后的值，也就是成功概率和失败概率的比值的自然对数，它可以用以下公式表示：

$$z_i = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$$

其中$\beta_0, \beta_1, \cdots, \beta_k$是回归系数，$x_1, x_2, \cdots, x_k$是自变量。

2. 请结合公式说明为什么logistic regression可以认为是某种形式的线性模型

Logistic regression可以认为是某种形式的线性模型，因为它的输出变量（即响应事件的概率）的对数几率（logit）是输入变量和参数的线性函数。

也就是说，logistic regression模型可以写成如下形式：

$$\operatorname{logit}(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \cdots + \beta_p x_{p,i}.$$

其中，p_i是第i个观测值的响应事件（例如0或1）发生的概率，$x_{j,i}$是第i个观测值的第j个输入变量，β_j是第j个参数。

这种形式与一般线性模型（generalized linear model）相同，只不过响应变量经过了一个非线性的转换函数（即对数几率函数）。

3. 请从最大似然估计maximum likelihood estimation (MLE)的角度推导出logistic regression的梯度上升的求解方法

logistic regression的梯度上升的求解方法可以从最大似然估计的角度推导出来。具体步骤如下：

假设我们有n个样本$(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$，其中$y_i$是二元分类标签，取值为0或1。我们假设每个样本的标签服从一个伯努利分布，即：

$$P(y_i|\mathbf{x}_i, \mathbf{w}) = \hat{p}_i^{y_i}(1-\hat{p}_i)^{1-y_i}$$

其中$\hat{p}_i$是logistic regression模型对样本$\mathbf{x}_i$属于类别1的概率预测，即：

$$\hat{p}_i = \frac{1}{1+e^{-\mathbf{w}^T\mathbf{x}_i}}$$

其中$\mathbf{w}$是模型参数向量。

那么，最大似然估计就是要找到一组参数$\mathbf{w}$，使得所有样本标签出现的联合概率最大化，即：

$$L(\mathbf{w}) = \prod_{i=1}^{n} P(y_i|\mathbf{x}_i, \mathbf{w})$$

为了方便计算，我们通常对上式取对数，并且加上一个负号变成最小化问题，即：

$$J(\mathbf{w}) = -\log L(\mathbf{w}) = -\sum_{i=1}^{n}[y_i \log \hat{p}_i + (1-y_i)\log(1-\hat{p}_i)]$$

这就是交叉熵损失函数（cross-entropy loss function）。

为了求解这个优化问题，我们可以使用梯度下降法（gradient descent），也就是不断地沿着损失函数的负梯度方向更新参数。而梯度上升法（gradient ascent）则是沿着损失函数的正梯度方向更新参数。两者只有符号上的区别。

那么，我们需要计算损失函数$J(\mathbf{w})$关于参数$\mathbf{w}$的梯度。根据链式法则和导数公式，我们有：

$$\nabla J(\mathbf{w}) = -\sum_{i=1}^{n}[y_i(1-\hat{p}_i) - (1-y_i)\hat{p}_i] \cdot \nabla\left(\frac{-e^{-\mathbf{x}_i^T\cdot\boldsymbol{\beta}}}{(1+e^{-\boldsymbol{\beta}\cdot x_i})^2}\right)$$
$$= -(-e^{-x_{ij}\beta_j})(-x_{ij})(-e^{-x_{ij}\beta_j})(-x_{ij})(-e^{-x_{ij}\beta_j})(-x_{ij}) = x_{ij}(y_i - \hat{p}_i)$$

所以，使用梯度上升法求解最大似然估计问题时，每次迭代时需要更新参数为：

$$\boldsymbol{\beta}_{j+1} = \boldsymbol{\beta}_j + \alpha x_{ij}(y_i - \hat{p}_i)$$

其中j表示迭代次数，α表示学习率。