

LS-PLM模型 (MLR)

针对问题：LR表达能力差（阿里巴巴主流推荐模型）

解决思路：

- 通过聚类，构造【带权】LR
- 先对样本进行分片（聚类），再在样本分片中应用逻辑回归进行CTR评估

$$f(x) = \sum_{i=1}^m \pi_i(x) \cdot \eta_i(x) = \sum_{i=1}^m \frac{e^{\mu_i \cdot x}}{\sum_{j=1}^m e^{\mu_j \cdot x}} \cdot \frac{1}{1 + e^{-w_i \cdot x}}$$

聚类函数
这里用的是softmax
逻辑回归

其中m代表分片数，当m=1时退化为LR模型，m越大拟合能力越强规模也就越大，阿里给出的m经验值为12.

目标函数

目标函数如公式(2)所示

$$\arg \min_{\Theta} f(\Theta) = \text{loss}(\Theta) + \lambda \|\Theta\|_{2,1} + \beta \|\Theta\|_1 \quad (2)$$

$\text{loss}(\Theta)$ 根据不同场景下不同，比如二分类用交叉熵作为损失函数：

$$\text{loss}(\Theta) = - \sum_{t=1}^n [y_t \log(p(y_t=1|x_t, \Theta)) + (1-y_t) \log(p(y_t=0|x_t, \Theta))]$$

后面为正则项，其中

$\|\Theta\|_1 = \sum_{ij} |\theta_{ij}|$ 是对每个参数的L1正则，保证所有参数的稀疏性；

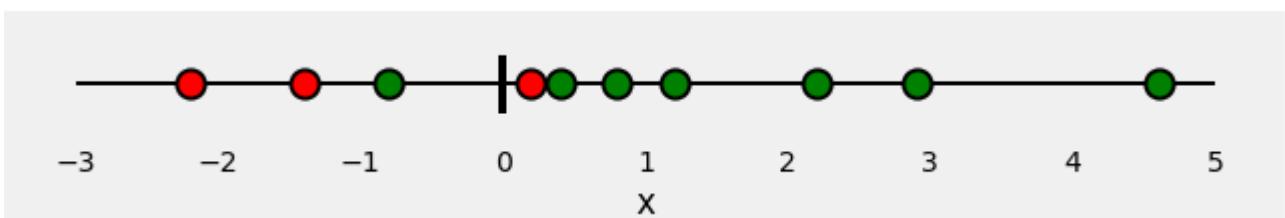
$\|\Theta\|_{2,1} = \sum_{i=1}^d \sqrt{\sum_{j=1}^{2m} \theta_{ij}^2}$ 是对L2正则的L1正则，根号里面是对某个特征的2m个参数的L2正则，外面是L1，这样是为了保证特征的稀疏性，做feature selection。

正则部分相当于对整体参数和按照特征分组的参数分别做了正则，既在最细粒度上筛选参数也在较粗的粒度上筛选特征。

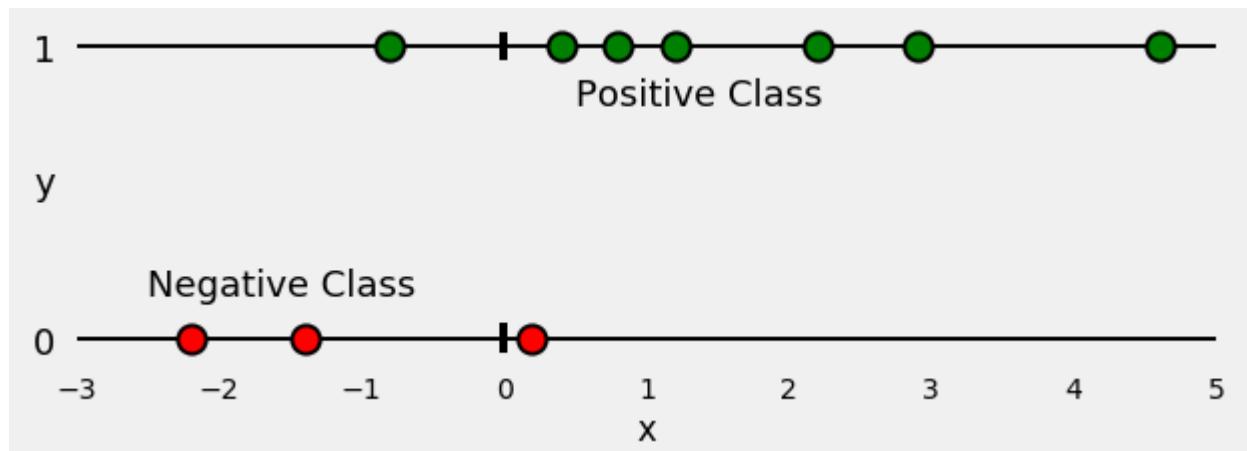
补充：

二元交叉熵

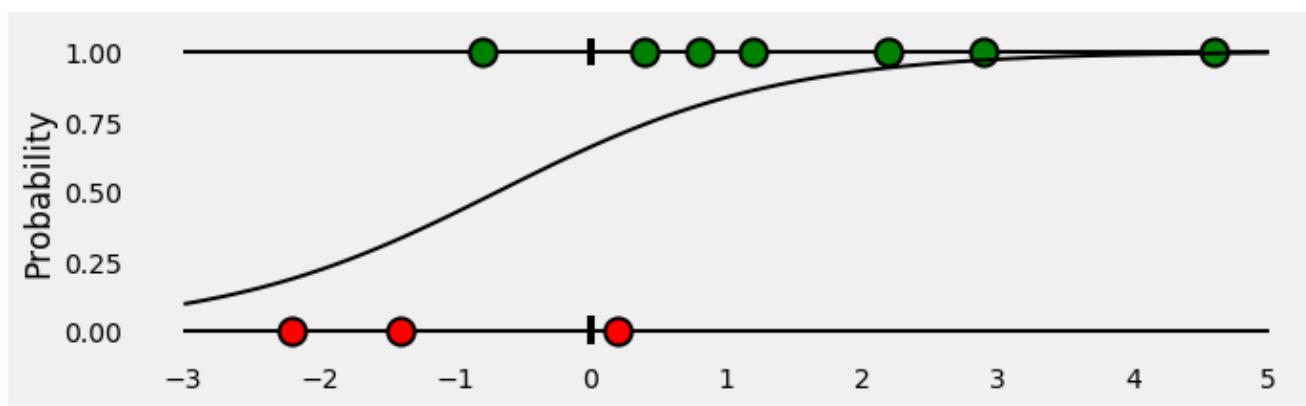
$$H_p(q) = -\frac{1}{N} \sum_{i=1}^N y_i \cdot \log(p(y_i)) + (1-y_i) \cdot \log(1-p(y_i))$$



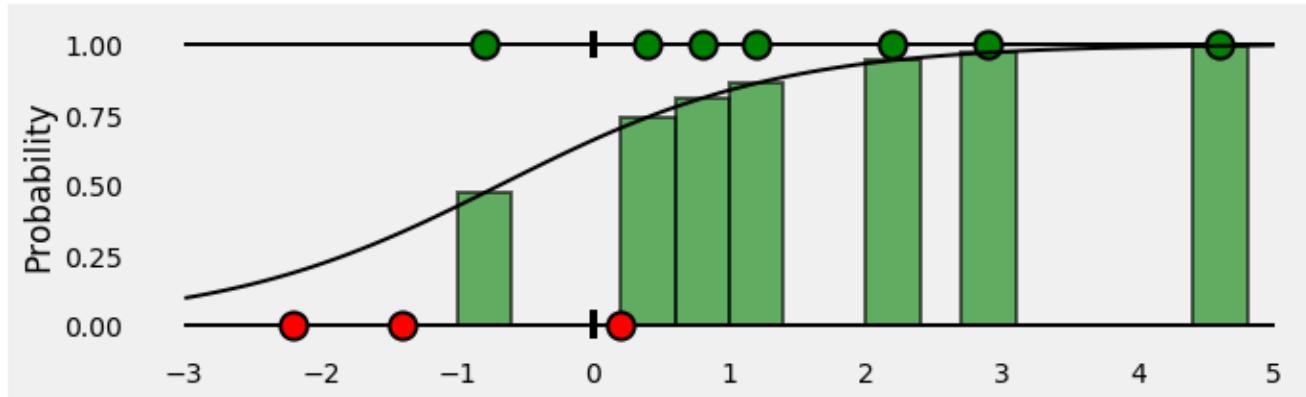
y是标签（1是绿色的，0是红色的），p(y)是所有的N个点预测是绿色的概率。



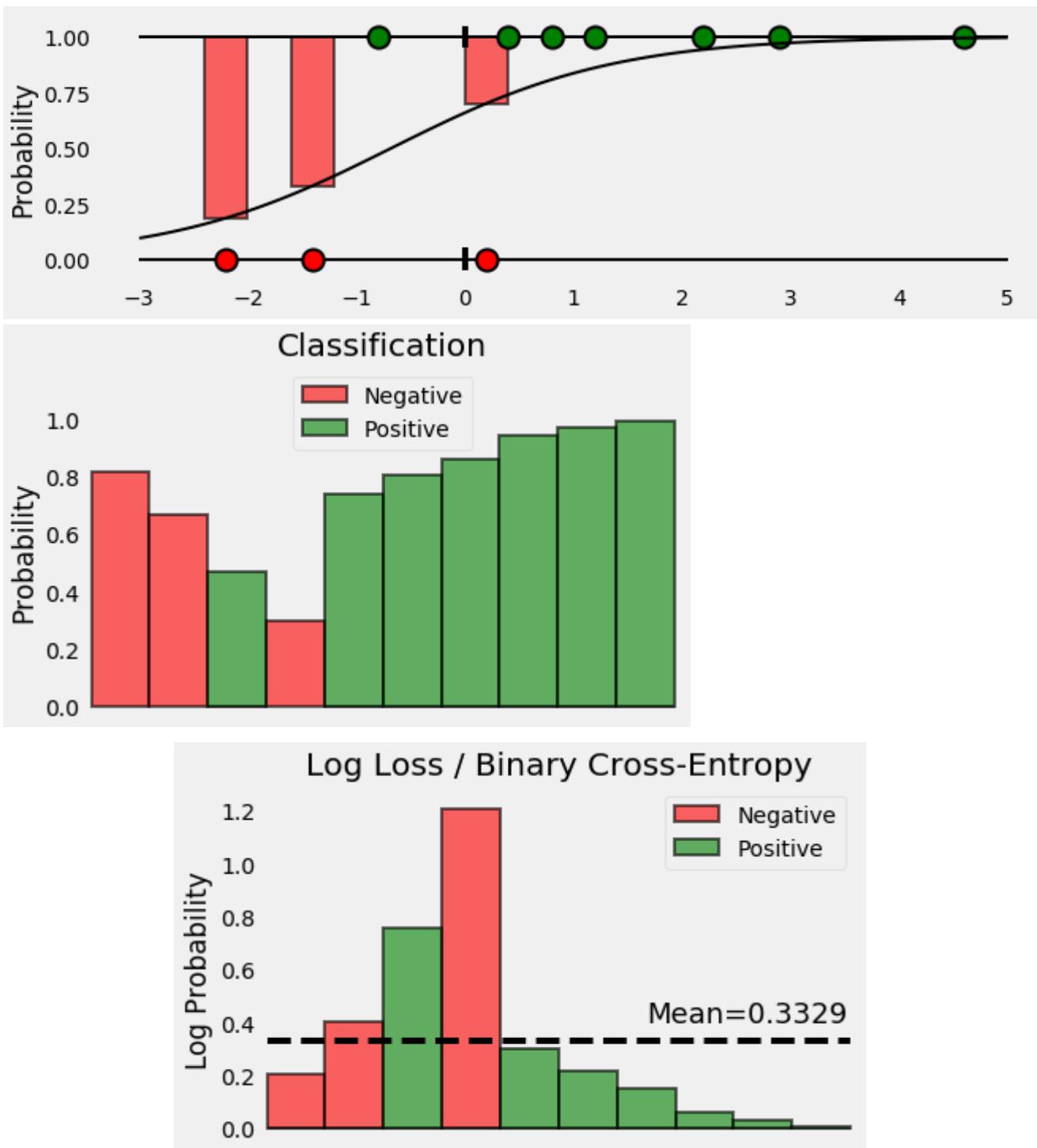
现在，我们来训练逻辑回归模型来分类我们的点。这个回归的拟合是一个sigmoid的曲线，表示了给定的 x 是绿色的概率。就像这样：



对于所有的属于正样本的点（绿色），我们的分类器给出的预测概率是什么？就是sigmoid曲线下面的绿色的条， x 的坐标代表了这个点。



到现在为止，一切都好！那么负样本的点呢？记住，sigmoid曲线之下的绿条表示的该点是绿色的概率。那么，给定的点是红色的概率是多少呢？当然就是sigmoid曲线上面红色条



L1L2L2,1正则：

L1正则化和L2正则化的说明如下：

- L1正则化是指权值向量 ww 中各个元素的绝对值之和
- L2正则化是指权值向量 ww 中各个元素的平方和然后再求平方根

$$X = [x_1 \ x_2 \ \dots \ x_d]$$

$$U_m = [u_{1m} \ u_{2m} \ \dots \ u_{dm}]$$

$$W_m = [w_{1m} \ w_{2m} \ \dots \ w_{dm}]$$

$$f(x) = \sum_{i=1}^m \pi_i(x) \cdot \eta_i(x) = \sum_{i=1}^m \frac{e^{\mu_i \cdot x}}{\sum_{j=1}^m e^{\mu_j \cdot x}} \cdot \frac{1}{1 + e^{-w_i \cdot x}}$$

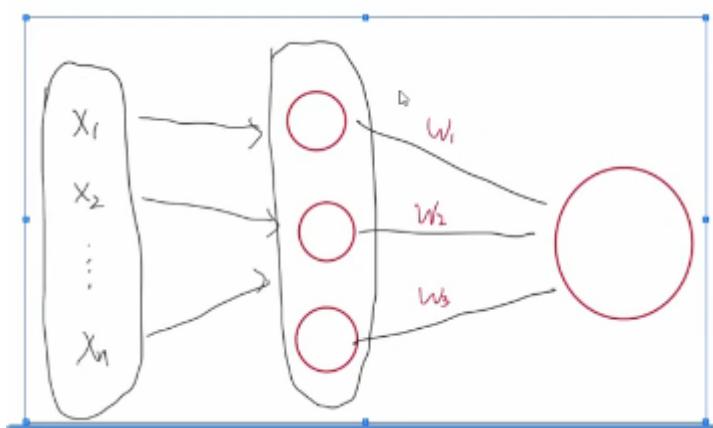
$$\begin{matrix} & m=1 & m=2 & \dots & m=m & m=1 & m=2 & \dots & m=M \\ X_1 & \left[\begin{matrix} U_{11} & U_{12} & \dots & U_{1m} & W_{11} & W_{12} & \dots & W_{1m} \\ \vdots & & & & & & & \\ U_{d1} & U_{d2} & \dots & U_{dm} & W_{d1} & W_{d2} & \dots & W_{dm} \end{matrix} \right] \\ \vdots \\ X_d & \end{matrix}$$

$$f(x) = \sum_{i=1}^m \pi_i(x) \cdot \eta_i(x) = \sum_{i=1}^m \frac{e^{\mu_i \cdot x}}{\sum_{j=1}^m e^{\mu_j \cdot x}} \cdot \frac{1}{1 + e^{-w_i \cdot x}}$$



聚类函数
这里用的是softmax

逻辑回归



MLR可以看做是加入了注意力的三层神经网络，输入层是样本的特征向量，中间层是m个神经元组成的隐层，其中m是分片的个数，最后一层是单个神经元的输出层。

问题：

分片之后是对不同区域的样本分别进行逻辑回归吗？