

## 第3章 线性模型2

---

### Exercise 3.7

如果类别标签为+1和-1，推导logistic回归的对数似然函数为：

$$-\sum_{i=1}^N \log(1 + \exp(-y_i(w^T \mathbf{x}_i + b)))$$

### Exercise 3.8

For logistic regression,  $y$  belongs to  $\{-1, 1\}$ , show that

$$\nabla E(w) = -\frac{1}{N} \sum_{n=1}^N \frac{y_n x_n}{1 + e^{y_n w^T x_n}} = \frac{1}{N} \sum_{n=1}^N -y_n x_n \theta(y_n w^T x_n)$$

Argue that a 'misclassified' example contributes more to the gradient than a correctly classified one.

注：目标函数使用的是对数似然函数

### Exercise 3.9

现在有一个数据集T,  $x = \{(1, 3), \{2, 2\}, \{3, 7\}, \{-2, 0\}, \{-3, -9\}\}$ , 其中对应的 $y$ 为 $\{1, 1, 1, -1, -1\}$ ，如果用感知机算法训练，那么其在T数据集上的误分类次数 $k$ 最多为？

## 实践题

---

### Exercise 3.2

题目介绍：

不知道你是否在朋友圈被刷屏过 NBA 的某场比赛进度或者结果？或者你就是一个 NBA 狂热粉，比赛中的每个进球、抢断或是逆转压哨球都能让你热血沸腾。除去观赏精彩的比赛过程，我们也同样好奇比赛的结果会是如何。因此本节课程，将给同学们展示如何使用 NBA 比赛的以往统计数据，判断每个球队的战斗力，及预测某场比赛中的结果。

我们将基于 2015-2016 年的 **NBA** 常规赛及季后赛的比赛统计数据，预测在当下正在进行的 2016-2017 常规赛每场赛事的结果（赢还是输）。

数据集介绍：

1、data1.csv(每支队伍平均每场比赛的表现统计)

数据名	含义
Rk -- Rank	排名
G -- Games	参与的比赛场数（都为 82 场）
MP -- Minutes Played	平均每场比赛进行的时间
FG--Field Goals	投球命中次数
FGA--Field Goal Attempts	投射次数
FG%--Field Goal Percentage	投球命中次数
3P--3-Point Field Goals	三分球命中次数
3PA--3-Point Field Goal Attempts	三分球投射次数
3P%--3-Point Field Goal Percentage	三分球命中率
2P--2-Point Field Goals	二点球命中次数
2PA--2-point Field Goal Attempts	二点球投射次数
2P%--2-Point Field Goal Percentage	二点球命中率
FT--Free Throws	罚球命中次数
FTA--Free Throw Attempts	罚球投射次数
FT%--Free Throw Percentage	罚球命中率
ORB--Offensive Rebounds	进攻篮板球
DRB--Defensive Rebounds	防守篮板球
TRB--Total Rebounds	篮板球总数
AST--Assists	助攻
STL--Steals	抢断
BLK -- Blocks	封盖
TOV -- Turnovers	失误
PF -- Personal Fouls	个犯
PTS -- Points	得分

## 2、data2.csv（对局信息）

主要有以下字段：

date: 比赛日期

start: 开始时间

vistor: 客队

home: 主队

PTS: 得分

3.data3.csv (需要预测结果)

含有date、start、vistor、home

数据集地址: <https://github.com/jjw12345/logistic/tree/main>

要求: 根据, 15年球队的信息以及对局结果, 试着使用逻辑回归学习一个二分类问题 (预测主客的输赢), 最后提交代码和16年的预测结果。

