

实验1总结

- 实验报告要点提示

- 算法原理：不抄PPT，更不能抄网上内容。要用**自己的话**，把**原理讲明白**（结合理论课所学知识，不需要讲实现层面的细节），有自己的见解更好。**不要太简略**（想象读者是“请假”的同学）。
- 伪代码或流程图：**关键步骤**不遗漏，注意**详略**，注意**规范**（可以复习一下“数据结构与算法”）。
- 关键代码：只写你认为关键的部分（main函数、文件IO等并不关键），注意**分块**与**说明**（代码注释&报告中的文字）。
- 实验结果：完整充分，**分析**到位（不同参数/设置下的实验结果的展示&现象描述&**原因分析**），**展示**清晰（尽量用可视化图/表）。
- 思考题：如果有多个小问，则都需要答。
- 创新：如有，要清晰列出，**讲清楚**，并进行**实验**对比分析。

- 引用未标注涉嫌抄袭



感知机 & 逻辑回归

张雪雯

2021/9/23



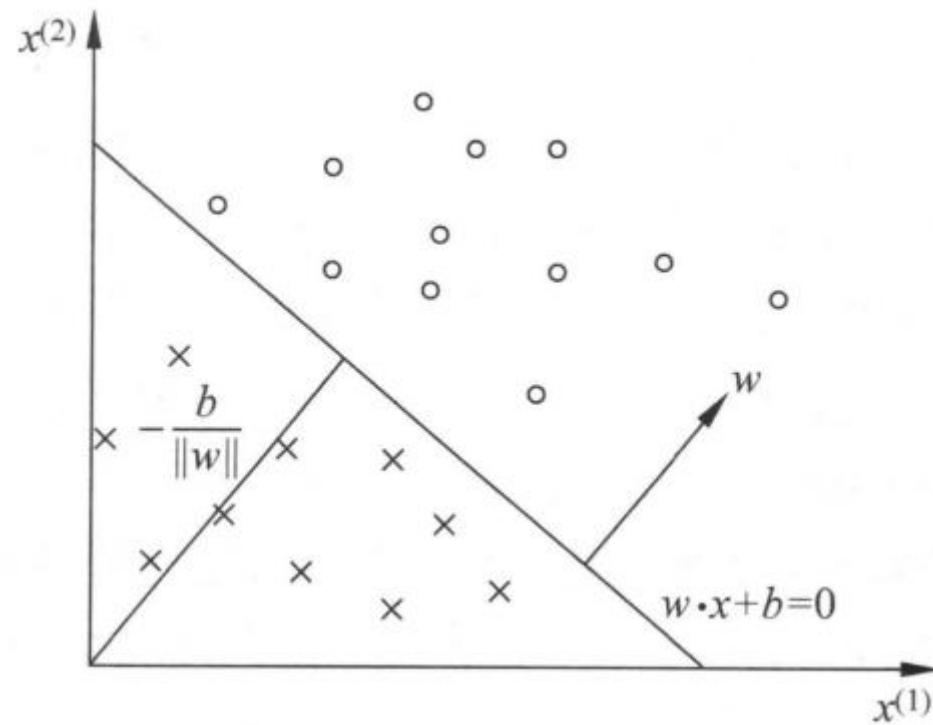
感知机 (PLA)

- 感知机针对二分类问题，输入是样本的特征向量 $x \in \mathbf{R}^n$ ，输出是样本的类别 $y \in \{+1, -1\}$
- 感知机可以表示为： $f(x) = \mathbf{sign}(w \cdot x + b)$
 - $\mathbf{sign}(x)$ 表示符号函数
 - w 和 b 表示模型参数， $w \in \mathbf{R}^n$ 称为权值向量， $b \in \mathbf{R}$ 称为偏置
- 感知机的几何解释：线性方程 $w \cdot x + b = 0$ 对应于特征空间 \mathbf{R}^n 的一个分离超平面；平面将特征空间分为两部分，位于两部分的点分别对应于正、负样本



感知机 (PLA)

- 感知机的几何解释





感知机 (PLA)

- 损失函数：所有误分类点到分离超平面的距离之和，感知机的学习目标是令该和尽可能小，即令误分类点的数量尽可能少
- 误分类点 (x_i, y_i) 到分离超平面的距离：
$$\frac{1}{\|w\|} \|w \cdot x_i + b\|$$
 - 由于 $-y_i(w \cdot x_i + b) > 0$ ，该距离可写成：
$$-\frac{1}{\|w\|} y_i (w \cdot x_i + b)$$
 - 假设误分类点集合为 M ，并且不考虑 $-\frac{1}{\|w\|}$ ，损失函数可写成：

$$L(w, b) = -\sum_{x_i \in M} y_i (w \cdot x_i + b)$$



感知机 (PLA)

- 使用随机梯度下降对损失函数进行优化

$$L(w, b) = -\sum_{x_i \in M} y_i (w \cdot x_i + b)$$

- 损失函数的梯度为:

$$\nabla_w L(w, b) = -\sum_{x_i \in M} y_i x_i$$

$$\nabla_b L(w, b) = -\sum_{x_i \in M} y_i$$

- 选择一个误分类点 (x_i, y_i) 对参数进行更新, 其中 η 表示学习率:

$$w = w + \eta y_i x_i$$

$$b = b + \eta y_i$$



感知机 (PLA)

- 感知机算法步骤:

1. 设置学习率 η ; 随机初始化 w 和 b , 一般可初始化为0
2. 选取一个误分类点 (x_i, y_i) , 即如果 $y_i(w \cdot x_i + b) \leq 0$, 则对参数进行更新

$$w = w + \eta y_i x_i$$

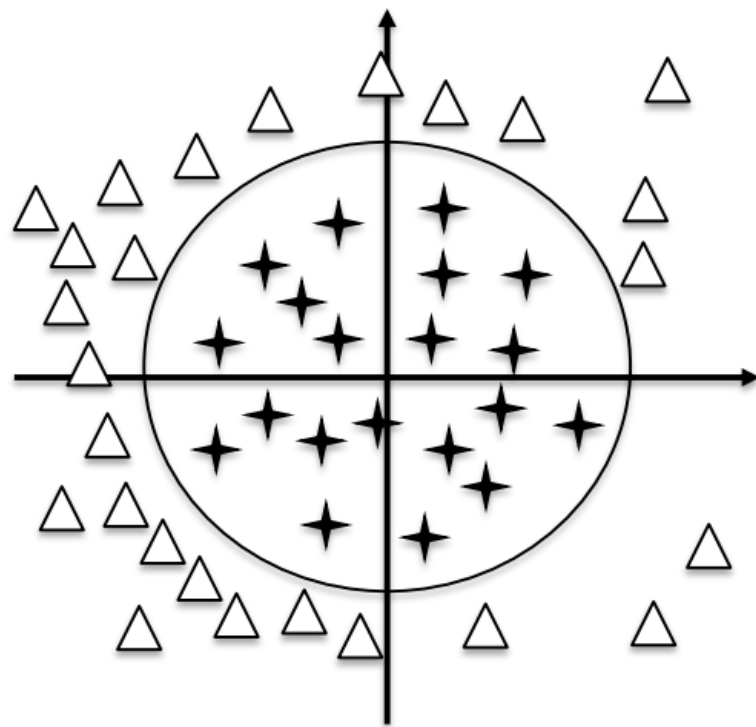
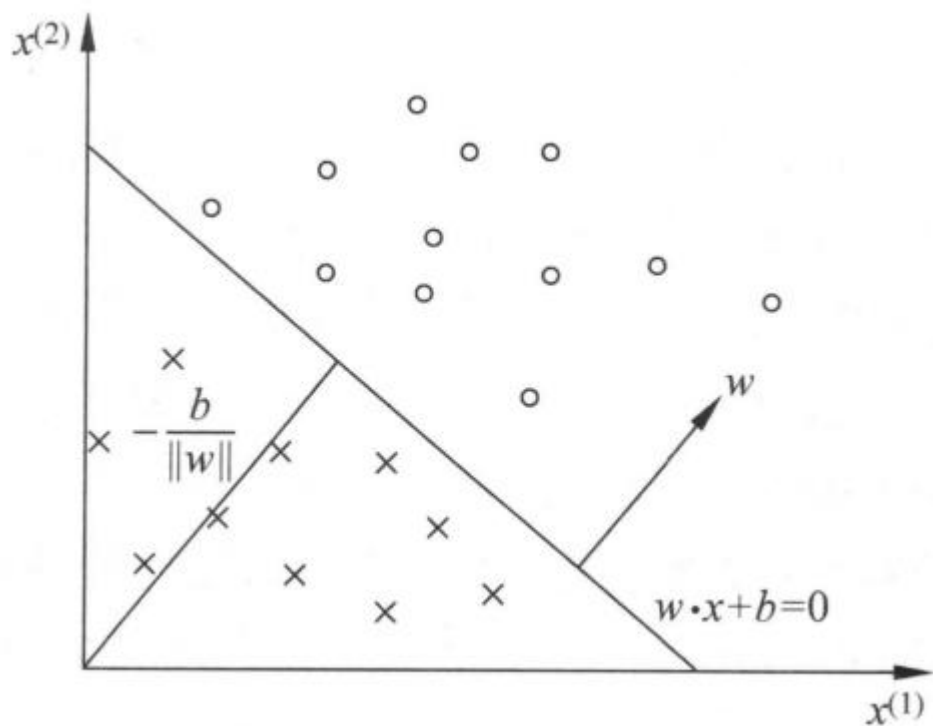
$$b = b + \eta y_i$$

3. 重复步骤2, 直到训练集中没有误分类点



感知机 (PLA)

- 线性可分与线性不可分





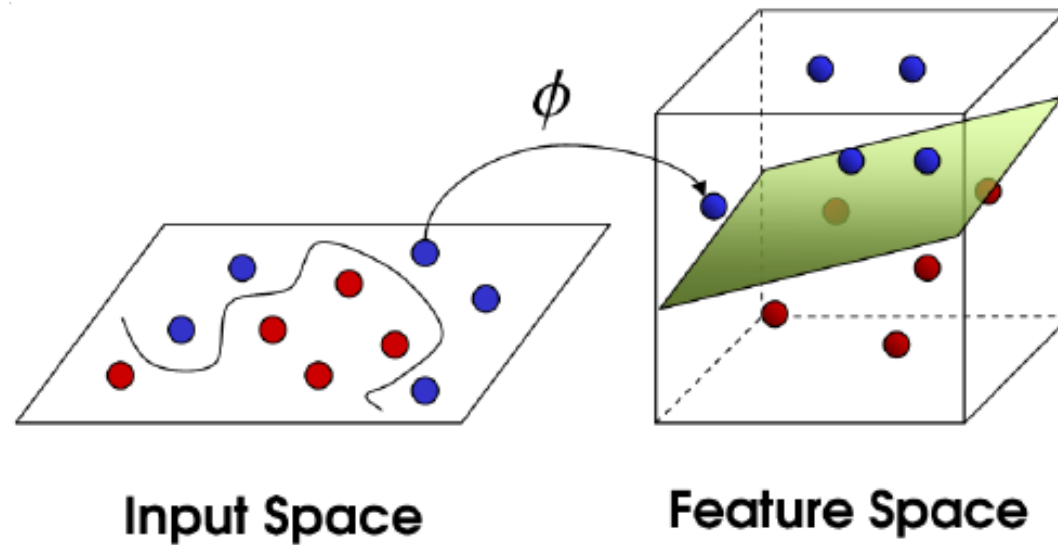
感知机 (PLA)

- 对于线性可分的数据集，总是可以找到一个模型将其正确划分
 - 线性可分：存在一个分离超平面，该平面能够将正、负样本完全正确地划分到平面的两侧
- 对于线性不可分的数据集，参考解决方案：
 - 设置最大迭代次数；当迭代次数到达预设值时，停止训练
 - 找到一组参数 (w, b) ，使得训练集使用该组参数进行划分后，分类错误的样本最少
 - 引入核函数 (kernel)



感知机 (PLA)

- 核函数 (kernel)



$$\phi(x) : \mathcal{X} \rightarrow \mathcal{H}$$



逻辑回归 (LR)

- 硬分类模型：非概率模型，通常表示为函数形式，即使用一个决策函数来直接判断样本的类别，如感知机、决策树等
- 软分类模型：概率模型，通常表示为概率分布形式，即先算出每个类别的概率，然后根据概率的大小来判断样本的类别，如逻辑回归



逻辑回归 (LR)

- 逻辑回归通常针对二分类问题，输入是样本的特征向量 $x \in \mathbf{R}^n$ ，输出是样本属于某个类别 $y \in \{0,1\}$ 的概率

- 逻辑回归可以表示为：

$$P(y=1|x) = \frac{\exp(w \cdot x + b)}{1 + \exp(w \cdot x + b)}, \quad P(y=0|x) = 1 - P(y=1|x)$$

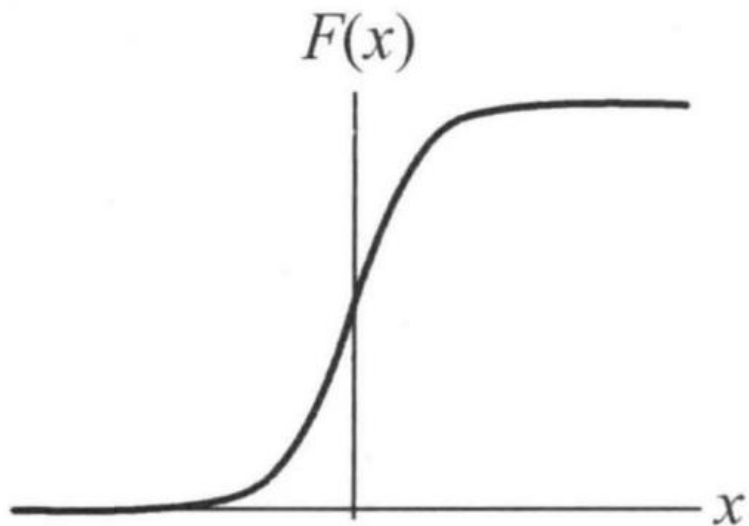
- 为方便表示，将 w 表示为 $w = (w^T, b)^T$ ，将 x 表示为 $x = (x^T, 1)^T$ ：

$$P(y=1|x) = \frac{\exp(w \cdot x)}{1 + \exp(w \cdot x)} = \frac{1}{1 + \exp(-w \cdot x)}, \quad P(y=0|x) = 1 - P(y=1|x)$$



逻辑回归 (LR)

- Logistic函数: $F(x) = \frac{1}{1 + \exp(-x)}$
 - $F(+\infty) = 1$, 当加权求和结果无穷大, 样本属于正类别的概率为1
 - $F(-\infty) = 0$, 当加权求和结果无穷小, 样本属于正类别的概率为0





逻辑回归 (LR)

- 令 $\pi(x) = \frac{1}{1 + \exp(-w \cdot x)}$, 则某个样本 (x, y) 预测为类别 y 的概率表示为:

$$f(x) = P(y | x) = \pi(x)^y (1 - \pi(x))^{1-y}$$

- 当 $y = 1$, $f(x) = P(y = 1 | x) = \pi(x)$
- 当 $y = 0$, $f(x) = P(y = 0 | x) = 1 - \pi(x)$
- $f(x)$ 称为似然函数



逻辑回归 (LR)

- 在整个训练集上考虑似然函数：

$$\prod_{i=1}^N [\pi(x_i)]^{y_i} [1 - \pi(x_i)]^{1-y_i}$$

- 对数似然函数为（似然函数取对数）：

$$\begin{aligned} L(w) &= \sum_{i=1}^N [y_i \log \pi(x_i) + (1 - y_i) \log(1 - \pi(x_i))] \\ &= \sum_{i=1}^N \left[y_i \log \frac{\pi(x_i)}{1 - \pi(x_i)} + \log(1 - \pi(x_i)) \right] \\ &= \sum_{i=1}^N [y_i (w \cdot x_i) - \log(1 + \exp(w \cdot x_i))] \end{aligned}$$



逻辑回归 (LR)

- 根据极大似然估计, 对 $L(w)$ 求极大值, 可得到参数 w 的估计值
- 对 $L(w)$ 取负, 将 $-L(w)$ 作为逻辑回归模型的损失函数, 并使用批量梯度下降法对损失函数进行优化

- 损失函数的梯度为:

$$-\nabla_w L(w) = -\sum_{i=1}^N [y_i - \pi(x_i)] x_i$$

- 对参数进行更新:

$$w = w + \eta \sum_{i=1}^N [y_i - \pi(x_i)] x_i$$



逻辑回归 (LR)

- 逻辑回归算法步骤:

1. 给每个样本的特征向量添加一维常数项1, 即 $x = (x^T, 1)^T$
2. 设置学习率 η ; 对n+1维的权值向量 $w = (w^T, b)^T$ 进行随机初始化
3. 计算当前梯度, 并对参数进行更新: $w = w + \eta \sum_{i=1}^N [y_i - \pi(x_i)] x_i$
4. 重复步骤3, 直至满足一定的收敛条件



思考题

- 随机梯度下降与批量梯度下降各自的优缺点？
- 不同的学习率 η 对模型收敛有何影响？从收敛速度和是否收敛两方面来回答。
- 使用梯度的模长是否为零作为梯度下降的收敛终止条件是否合适，为什么？一般如何判断模型收敛？



实验注意事项

- 实现感知机和基于批量梯度下降的逻辑回归，分别提交一份代码
- 本次的实验数据是train.csv，前40列表示特征，最后一列表示标签（0或1）
- 请自行分好训练集、测试集（在报告里说明怎么分的），评测指标为测试集上的准确率
- DDL:
 1. 验收：10.14周四 实验课
 2. 实验报告：10.13周三 23:00