机器学习导论习题二

学号, 姓名, 邮箱

2024 年 4 月 10 日

作业注意事项

1. 作业所需的 LaTeX 及 Python 环境配置要求请参考: [Link];
2. 请在 LaTeX 模板中第一页填写个人的学号、姓名、邮箱;
3. 本次作业需提交的文件与对应的命名方式为:
   (a) 作答后的 LaTeX 代码 — `HW2.tex`;
   (b) 由 (a) 编译得到的 PDF 文件 — `HW2.pdf`;
   (c) 编程题代码 — `main.py`;
   请将以上文件**打包为 "学号_姓名.zip"** (例如 "221300001_张三.zip") 后提交;
4. 若多次提交作业, 则在命名 ".zip" 文件时加上版本号, 例如 "221300001_张三_v1.zip" (批改时以版本号最高, 提交时间最新的文件为准);
5. 本次作业提交截止时间为 **4 月 19 日 23:59:59**. 未按照要求提交作业, 提交作业格式不正确, **作业命名不规范**, 将会被扣除部分作业分数; 除特殊原因 (如因病缓交, 需出示医院假条) 逾期未交作业, 本次作业记 0 分; **如发现抄袭, 抄袭和被抄袭双方成绩全部取消**;
6. 学习过程中, 允许参考 ChatGPT 等生成式语言模型的生成结果, 但必须在可信的信息源处核实信息的真实性; **不允许直接使用模型的生成结果作为作业的回答内容**, 否则将视为作业非本人完成并取消成绩;
7. 证明题请给出关键证明步骤, 计算题请列出算式及中间结果, 否则不予计分; 撰写数学表达式请遵循符号约定.
8. 本次作业提交地址为 [Link], 请大家预留时间提前上交, 以防在临近截止日期时, 因网络等原因无法按时提交作业.

符号约定

[**线性代数变量符号**] 本次作业的部分题目涉及矩阵形式的数学推导, 请注意符号规范.
提示: 可以在导言区通过 `\def`, `\newcommand` 等命令自定义常用符号以简化编写, 例如 `\Real`, `\norm` 等.

- 矩阵请使用**粗体正体字母**, 例如 $\mathbf{A}, \mathbf{X}, \mathbf{O}$ (零矩阵);
- 向量请使用**粗体斜体字母**, 例如 $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}, \boldsymbol{x}, \boldsymbol{y}, \mathbf{0}$ (零向量);
- 标量 (以及离散变量) 请使用**斜体小写字母**, 例如 $a, b, c, \lambda, \mu, \sigma$;
- 操作符或者函数名请使用**正体有衬线体字母**, 例如 $\min, \max, \mathrm{Ent}, \mathrm{MI}$;
- 转置矩阵 (Transposed Matrix) 以及埃尔米特矩阵 (Hermitian Matrix) 的**角标**请使用**正体无衬线体字母**, 例如 $\mathbf{A}^{\mathsf{T}}, \mathbf{A}^{\mathsf{H}}$.

[**奇异值分解与特征值分解**] 由于 $\mathbb{R}^n \times \mathbb{R}^m$ 上存在全序关系, 我们可以约定, 存在一种奇异值分解过程, 在此奇异值分解过程的意义下, 任何矩阵的奇异值分解是唯一的, 进而**任何 (半) 正定矩阵的特征值分解也是唯一的**. 如果没有符号冲突, 可以不加说明的使用以下记号:

- 矩阵 $\mathbf{A} \in \mathbb{R}^{n\times m}$ 的**奇异值分解**是 $\mathbf{A} = \sum_{i=1}^{\min\{m,n\}} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^{\mathsf{T}}$, $\sigma_i \geqslant \sigma_{i+1}$;
- **(半) 正定**矩阵 $\mathbf{A} \in \mathbb{R}^{n\times n}$ 的**特征值分解**是 $\mathbf{A} = \sum_{i=1}^{n} \lambda_i \boldsymbol{v}_i \boldsymbol{v}_i^{\mathsf{T}}$, $\lambda_i \geqslant \lambda_{i+1}$;
- 另记 $\sigma_{\max}(\mathbf{A}) = \sigma_1(\mathbf{A}), \lambda_{\max}(\mathbf{A}) = \lambda_1(\mathbf{A})$ 是矩阵 $\mathbf{A}$ 最大的奇异值.

[**决策树**] 为简化表述, 我们引入指示函数 $\mathbb{I}(e)$: 若事件 e 发生, 则 $\mathbb{I}(e) = 1$, 否则 $\mathbb{I}(e) = 0$. 对于**表格数据**, 我们约定:

- 每个样例 $\boldsymbol{x}$ 都具有 d 个属性 $\{1, \cdots, d\}$;
- 第 j 个属性有 n_j 种可能的取值, 记作 $\boldsymbol{x}^j \in \{v_1^j, \cdots, v_{n_j}^j\}$;
- 对于多分类问题, 样例的标签可以记作 $\boldsymbol{x}^\ell$, 共有 N 种可能取值, 即 $\boldsymbol{x}^\ell \in \{1, \cdots, N\}$.

例如: 对于数据集 $D = \{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_m\}$, 事件 $\boldsymbol{x}^j = v_k^j$ 的频率值可以记作 $\hat{p}(\boldsymbol{x}^j = v_k^j) = \frac{1}{m}\sum_{i=1}^{m} \mathbb{I}\left(\boldsymbol{x}_i^j = v_k^j\right)$; 再如, 对于 "西瓜数据集 2.0" (详见教材表 4.1), 记号示范如下:

- $d = 6$;
- $\boldsymbol{x}^1 \in \{v_1^1 = \text{乌黑}, v_2^1 = \text{青绿}, v_3^1 = \text{浅白}\}$, 即第 1 个属性色泽, $n_1 = 3$;
- $\boldsymbol{x}^2 \in \{v_1^2 = \text{蜡缩}, v_2^2 = \text{稍蜷}, v_3^2 = \text{硬挺}\}$, 即第 2 个属性根蒂, $n_2 = 3$;
- $\boldsymbol{x}^3 \in \{v_1^3 = \text{清脆}, v_2^3 = \text{浊响}, v_3^3 = \text{沉闷}\}$, 即第 3 个属性敲声, $n_3 = 3$;
- $\boldsymbol{x}^4 \in \{v_1^4 = \text{清晰}, v_2^4 = \text{稍糊}, v_3^4 = \text{模糊}\}$, 即第 4 个属性纹理, $n_4 = 3$;
- $\boldsymbol{x}^5 \in \{v_1^5 = \text{凹陷}, v_2^5 = \text{稍凹}, v_3^5 = \text{平坦}\}$, 即第 5 个属性脐部, $n_5 = 3$;
- $\boldsymbol{x}^6 \in \{v_1^6 = \text{软粘}, v_2^6 = \text{硬滑}\}$, 即第 6 个属性触感, $n_6 = 2$.

1 [20pts] 岭回归

在本题中, 我们假设**所有凸优化问题均有解**.
回顾教材第三章第二节, 多元线性回归相当于求解如下的无约束优化问题 (1.1). 其中, 对于 $\boldsymbol{v} \in \mathbb{R}^d$, 定义 $\|\boldsymbol{v}\|_2^2 = \boldsymbol{v}^\mathsf{T}\boldsymbol{v}$.

$$\min_{\hat{\boldsymbol{w}}} \quad \|\mathbf{X}\hat{\boldsymbol{w}} - \boldsymbol{y}\|_2^2 \tag{1.1}$$

在多元线性回归中增加约束项, 即成为**岭回归 (Rigde Regression)**, 即求解有约束优化问题 (1.2), 其中 $\rho > 0$ 是待确定的超参数.

$$\begin{aligned} \min_{\hat{\boldsymbol{w}}} \quad & \|\mathbf{X}\hat{\boldsymbol{w}} - \boldsymbol{y}\|_2^2 \\ \text{s.t.} \quad & \|\hat{\boldsymbol{w}}\|_2^2 \leqslant \rho^2 \end{aligned} \tag{1.2}$$

岭回归也可以写成无约束优化问题 (1.3). 其中, $\lambda \geqslant 0$ 是待确定的超参数.

$$\min_{\hat{\boldsymbol{w}}} \quad \|\mathbf{X}\hat{\boldsymbol{w}} - \boldsymbol{y}\|_2^2 + \lambda \|\hat{\boldsymbol{w}}\|_2^2 \tag{1.3}$$

(1) **[5pts]** 相比于多元线性回归, 岭回归引入了额外的**归纳偏好 (Inductive Bias)**. 回顾教材第一章第四节, 请简要回答: 岭回归引入了怎样的归纳偏好? 这样的归纳偏好有怎样的作用?
提示: 回顾**过拟合 (Overfitting)**、**"奥卡姆剃刀" (Occam's Razor) 原则**等知识点; 结合特殊情形回答, 例如矩阵 $\mathbf{X}$ 不满秩、数据中存在异常值等.

(2) **[5pts]** 请证明岭回归的两种形式 (1.2) 和 (1.3) 等价.
提示: 考虑 **KKT 条件 (Karush-Kuhn-Tucker Conditions)**.

(3) **[5pts]** 对于无约束优化形式 (1.3), 假设 λ 已经确定, 此时岭回归的解记作 $\boldsymbol{w}^\star$, 请推导出 $\boldsymbol{w}^\star$ 的表达式.

(4) **[5pts]** **在 (3) 的基础上**, 请推导出 λ 的下界 (关于 ρ 的函数), 并据此回答: ρ 减小时, 若希望保持 (1.2) 和 (1.3) 的解一致, 需要怎样调整 λ?
提示: 你可能需要引入 $\sigma_{\max}(\mathbf{X})$.

Solution. 此处用于写解答 (中英文均可)

2 [20pts] 决策树的构建流程

> **[注意事项]** 本题可使用 PowerPoint®, Visio® 等软件软件绘制决策树, 导出为图片或 PDF 插入到作业文档中; 亦允许手绘决策树, 但是请确保文字与线条的工整. 如果插入照片, 请确保照片内容清晰可读, 否则将扣除部分分数.

考虑如下的表格数据集: 在诊断疾病 RD 时, 通常采用 UW-OCT, OCT-PD, US 三种检测手段, 其中 1 代表阳性, 0 代表阴性. 假设总共收集了 16 位病人的检测结果, 其中 8 条用于训练, 如表格 1 所示, 8 条用于验证, 如表格 2 所示.

表 1: 训练集 D_{train}

编号	UW-OCT	OCT-PD	US	RD
1	1	1	0	1
2	1	1	1	1
3	0	1	1	1
4	0	1	1	1
5	0	0	1	0
6	1	0	0	0
7	1	0	1	0
8	0	1	0	0

表 2: 验证集 D_{valid}

编号	UW-OCT	OCT-PD	US	RD
9	1	1	1	1
10	0	1	1	1
11	0	1	0	0
12	1	0	1	0
13	0	1	1	1
14	1	1	0	0
15	1	0	0	0
16	0	0	0	0

(1) **[10pts]** 回顾教材第四章第一, 二节, 请使用基尼指数作为划分准则, 通过训练集中的数据训练决策树. 在 `HW2.pdf` 中展示**最终得到的决策树**, 并给出**每个节点划分时的计算和比较过程**.

(2) **[5pts]** 回顾教材第四章第三节, 在 (1) 的基础上, 请判断每个节点是否需要预剪枝. 在 `HW2.pdf` 中展示**预剪枝后的决策树**, 并给出**每个节点预剪枝时的计算和比较过程**.

(3) **[5pts]** 对一个学习任务来说, 给定属性集, 其中有些属性可能很关键, 很有用, 称为 “相关特征”, 另一些属性则可能用处不大, 称为 “无关特征”. 请简要回答如下问题:

(a) 比较 (1,2) 的结果, 指出当前训练集和验证集划分下的无关特征, 并说明理由.

(b) 如果不给出数据集, 只给出决策树和剪枝后的决策树, 应该怎样挑选无关特征?

(c) 如果不给出数据集, 也不给出剪枝后的决策树, 只给出未剪枝的决策树, 还能挑选无关特征吗? 请简要给出理由.

Solution. 此处用于写解答 (中英文均可)

3 [15pts] 决策树的划分准则

(1) **[5pts]** 回顾教材第四章第一节, 请结合**决策树基本算法中递归返回的条件**, 简要回答: 如果要求 "只要训练集不含冲突数据 (即特征向量完全相同但标记不同), 就必须获得与训练集一致 (即训练误差为 0) 的决策树", 那么纯度函数需要满足怎样的要求?

(2) **[5pts]** 回顾教材第四章第二节, 信息增益可以重写成互信息 (Mutual Information)

$$\mathrm{MI}(E,F)=\sum_{e\in E}\sum_{f\in F}p(e,f)\log\frac{p(e,f)}{p(e)p(f)},$$

其中 E,F 都是事件的集合. 请给出此时 E,F 的定义, 列出必要的推导步骤, 并使用 $\mathrm{MI}(E,F)$, $\mathrm{Ent}(E)$, $\mathrm{Ent}(F)$ 等符号表示增益率.

(3) **[5pts]** 考虑归一化互信息 (Normalized Mutual Information) 的一种定义

$$\mathrm{NMI}(E,F)=\frac{\mathrm{MI}(E,F)}{\left(\mathrm{MI}(E,E)+\mathrm{MI}(F,F)\right)/2}=\frac{2\cdot\mathrm{MI}(E,F)}{\mathrm{Ent}(E)+\mathrm{Ent}(F)}.$$

在 (2) 的基础上, 如果使用归一化互信息作为划分准测, 与使用增益率作为划分准测产生的决策树相同吗? 请给出证明或举出反例.
提示: 已知数学性质 $0\leqslant\mathrm{MI}(E,F)\leqslant\min\{\mathrm{Ent}(E),\mathrm{Ent}(F)\}$.

Solution. 此处用于写解答 (中英文均可)

4 [20(+5)pts] 线性判别分析

回顾教材第三章第四节, LDA (Linear Discriminant Analysis) 有两个优化目标: 最小化类内散度 $\boldsymbol{w}^{\mathsf{T}}\mathbf{S}_{\mathrm{w}}\boldsymbol{w}$ 与最大化类间散度 $\boldsymbol{w}^{\mathsf{T}}\mathbf{S}_{\mathrm{b}}\boldsymbol{w}$, **目的是使得同类样例的投影点尽可能接近, 异类样例的投影点尽可能远离**. 在 LDA 之外, 课堂上还介绍了 PCA (Principal Components Analysis, 主成分分析). 事实上, PCA 可以写成类似 LDA 的形式, 但 PCA 只有一个目标, 即最大化全局散度: $\max_{\boldsymbol{w}} \boldsymbol{w}^{\mathsf{T}}\mathbf{S}_{\mathrm{t}}\boldsymbol{w}$.

(1) **[5pts]** 教材图 3.3 中, "+", "-" 代表数据点, 任务需要把二维数据降维到一维, 直线 $y = \boldsymbol{w}^{\mathsf{T}}\boldsymbol{x}$ 代表 LDA 的投影方向. 请画出图 1 数据集上 **PCA 的大致投影方向** (可以使用蓝色直线), 并在 `HW2.pdf` 中展示.

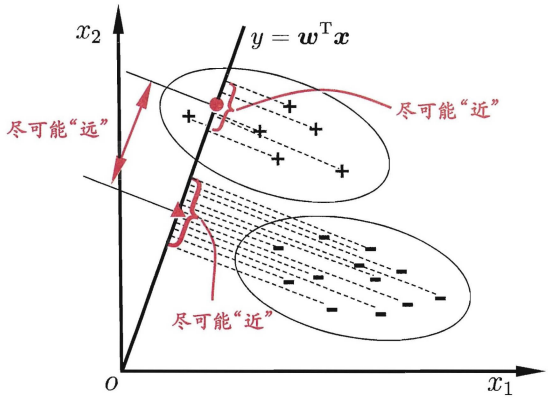

图 1: 教材图 3.3

(2) **[5pts]** 请参考题干中的介绍与 (1) 中的现象, 简要回答:

(a) 对照题干中 LDA 的优化目的, PCA 的优化目的是什么?

(b) PCA 相较于 LDA 有什么显著的不同点?

(3) **[5pts]** 下面, 我们先回顾教材第三章第四节中多分类 *LDA* 优化问题的矩阵形式. 考虑总类内散度是各个类别散度之和, 其矩阵形式为: $\mathbf{S}_{\mathrm{w}} = \sum_{i=1}^{N}\mathbf{S}_{\mathrm{w}i}$. 对于第 i 个类别的类内散度矩阵定义如下: $\mathbf{S}_{\mathrm{w}i} = \sum_{\boldsymbol{x}\in X_i}(\boldsymbol{x}-\boldsymbol{\mu}_i)(\boldsymbol{x}-\boldsymbol{\mu}_i)^{\mathsf{T}}$. 类似的, 总类间散度是各个类别中心相对于全局中心的散度之和, 其矩阵形式为: $\mathbf{S}_{\mathrm{b}} = \sum_{i=1}^{N}\mathbf{S}_{\mathrm{b}i}$. 对于第 i 个类别的中心相对于全局中心的散度矩阵定义如下: $\mathbf{S}_{\mathrm{b}i} = m_i(\boldsymbol{\mu}_i-\boldsymbol{\mu})(\boldsymbol{\mu}_i-\boldsymbol{\mu})^{\mathsf{T}}$.
LDA 事实上是在最小化**平均**类内散度和最大化**平均**类间散度, 其矩阵形式如 (4.1) 所示. 其中, d' 是降维后的维度, 严格小于数据维度 d.

$$\begin{aligned}\max_{\mathbf{W}} \quad & J(\mathbf{W}) = \frac{\mathrm{tr}\left(\mathbf{W}^{\mathsf{T}}\mathbf{S}_{\mathrm{b}}\mathbf{W}\right)}{\mathrm{tr}\left(\mathbf{W}^{\mathsf{T}}\mathbf{S}_{\mathrm{w}}\mathbf{W}\right)} = \frac{\mathrm{tr}\left(\mathbf{W}^{\mathsf{T}}\left(\sum_{i=1}^{N}\mathbf{S}_{\mathrm{b}i}\right)\mathbf{W}\right)}{\mathrm{tr}\left(\mathbf{W}^{\mathsf{T}}\left(\sum_{i=1}^{N}\mathbf{S}_{\mathrm{w}i}\right)\mathbf{W}\right)} = \frac{\frac{1}{N}\sum_{i=1}^{N}\mathrm{tr}\left(\mathbf{W}^{\mathsf{T}}\mathbf{S}_{\mathrm{b}i}\mathbf{W}\right)}{\frac{1}{N}\sum_{i=1}^{N}\mathrm{tr}\left(\mathbf{W}^{\mathsf{T}}\mathbf{S}_{\mathrm{w}i}\mathbf{W}\right)} \\ \text{s.t.} \quad & \mathbf{W}^{\mathsf{T}}\mathbf{W} = \mathbf{I}_{d'}\end{aligned} \tag{4.1}$$

根据教材中的介绍, (4.1) 可通过广义特征值分解进行求解. 然而, 在某些现实场景下, 我们应用 LDA 的目的是提高分类准确率, 那么通常进一步希望**每个**类别散度尽可能

小, **每个**类别中心相对于全局中心的散度尽可能大, **而非平均散度**. 因此, 考虑 LDA 的一种拓展形式:

$$\begin{aligned} \max_{\mathbf{W}} \quad & \left(\min_{i,j} J_{i,j}(\mathbf{W})\right) = \frac{\min_j\{\mathrm{tr}\left(\mathbf{W}^\mathsf{T}\mathbf{S}_{\mathrm{b}j}\mathbf{W}\right)\}}{\max_i\{\mathrm{tr}\left(\mathbf{W}^\mathsf{T}\mathbf{S}_{\mathrm{w}i}\mathbf{W}\right)\}} \\ \text{s.t.} \quad & \mathbf{W}^\mathsf{T}\mathbf{W} = \mathbf{I}_{d'} \end{aligned} \tag{4.2}$$

请指出拓展形式 (4.2) **无法直接沿用原有形式** (4.1) **的广义特征值求解算法的原因.**
提示: 指出求解时存在变量间的循环依赖关系.

(4) **[5pts]** *在线性代数中, 对于 (半) 正定矩阵* $\mathbf{A}, \mathbf{B}$*, 若* $(\mathbf{A}-\mathbf{B})$ *是正定矩阵, 则通常记作* $\mathbf{A} \succ \mathbf{B}$ *或* $\mathbf{B} \prec \mathbf{A}$*; 若* $(\mathbf{A}-\mathbf{B})$ *是半正定矩阵, 则通常记作* $\mathbf{A} \succcurlyeq \mathbf{B}$ *或* $\mathbf{B} \preccurlyeq \mathbf{A}$*. 在优化问题中, 凸优化问题有多项式时间复杂度的理论保证, 能够高效求解. 凸优化问题的定义是: (i) 最小化的目标函数是凸函数, 或者最大化的目标函数是凹函数, 而且 (ii) 可行域是凸集合. 可行域是所有满足约束条件的控制变量取值 (又称可行解) 构成的集合.*
拓展形式 (4.2) 不能沿用原有形式 (4.1) 的求解算法, 也不是凸优化问题. 为了高效求解, 需要探索一种将其转化成凸优化问题的方法. 已知原有形式 (4.1) 可以松弛成如下凸优化问题:

$$\begin{aligned} \max_{\mathbf{W},r} \quad & r \\ \text{s.t.} \quad & r \cdot \mathrm{tr}\left(\mathbf{S}_{\mathrm{w}}\mathbf{M}\right) - \mathrm{tr}\left(\mathbf{S}_{\mathrm{b}}\mathbf{M}\right) \leqslant 0 \\ & -r \leqslant 0 \\ & \mathbf{O} \preccurlyeq \mathbf{M} \preccurlyeq \mathbf{I}_d \\ & \mathrm{tr}\left(\mathbf{M}\right) = d' \end{aligned} \tag{4.3}$$

请仿照原有形式 (4.1) 的松弛形式 (4.3), 给出拓展形式 (4.2) 的松弛形式, 并证明拓展形式的松弛形式是凸优化问题, 即同时满足条件 (i) 和条件 (ii).

(5) **[5pts]** (*本题为附加题, 得分计入卷面分数, 但本次作业总得分不超过 100 分*) 请证明:

(a) 松弛形式 (4.1) 和原有形式 (4.3) 的约束条件不等价;

(b) 当 $r \cdot \mathrm{tr}\left(\mathbf{S}_{\mathrm{w}}\mathbf{M}\right) - \mathrm{tr}\left(\mathbf{S}_{\mathrm{b}}\mathbf{M}\right) = 0$ 时, (4.3) 的可行域是 (4.1) 可行域的凸包 (Convex Hull). 即: (4.3) 的可行解可以表示成 (4.1) 的可行解的线性组合.

进而, (4.3) 的可行域是包含 (4.1) 的可行域的最小凸集合, 即 (4.3) 对 (4.1) 的放松程度是最小的, 因而能够使得凸问题 (4.3) 的解尽可能的接近原问题 (4.1) 的解.

Solution. 此处用于写解答 (中英文均可)

5 [25pts] 编程实验: LDA 与多分类

[注意事项] 请不要修改或提交 `utils.py`; 在合理范围内, 运行时间和错误率不作为评分依据. 实现过程中只允许使用 NumPy 和 SciPy 提供的矩阵运算接口, 否则相应题目不计入分数.
此外, 对于题 (3,4,5), 如果调用 Sci-Kit Learn 实现二分类模型, 基于二分类模型实现多分类模型, 并且画出相应图像, 可计入题 (4,5) 得分, 但不计入题 (3) 得分.
[符号约定] `x` 是形状为 `(m, d)` 的矩阵, 其元素为 32 位浮点数; 在题 (1,3,4,5) 中, `y` 是形状为 `(m,)` 的向量, 其元素为 32 位整数; 在题 (2) 中, `y` 是形状为 `(m, 2)` 的向量, 其元素为 32 位浮点数. 其中: `m` 为样例个数, `d` 为样例维度; 标签从 `0` 开始, 例如共 20 类时, 标签依次是 $\{0, \cdots, 19\}$.

(1) **[5pts]** 根据 `main.py` 中的框架代码, 实现 LDA 降维, 通过 `lda_sanity_check` 测试, 并在 `HW2.pdf` 中展示运行后的终端输出截图.

(2) **[5pts]** 基于 (1) 分别把训练数据和测试数据降至两维, 并绘制在同一张散点图上, 在 `HW2.pdf` 中展示. 注意: 同类别的点应当使用同一颜色, 不同类别的数据点应当使用不同颜色.

(3) **[5pts]** 分类任务可以被归结为一种特殊的回归任务, 可参考 `sklearn` 中的内容: [Link]. 对于二分类任务, 我们任选一类作为正类, 另一类成为负类. 对于正类样本 $\boldsymbol{x}_+$, 约定 $y_+ = 1$, 对于负类样本 $\boldsymbol{x}_-$, 约定 $y_- = -1$, 对于训练得到的分类器 f 和测试样例 $\boldsymbol{x}$, 如果 $f(\boldsymbol{x}) \geqslant 0$ 预测为正类, 否则预测为负类.
根据框架代码, 按照上述约定实现基于岭回归的二分类模型, 通过 `classifier_2_sanity_check` 测试, 并在 `HW2.pdf` 中展示运行后的终端输出截图.

(4) **[5pts]** 基于 (3) 中的二分类模型, 通过 OvR 策略将其拓展为多分类模型, 通过 `classifier_n_sanity_check` 测试, 最后在 `HW2.pdf` 中展示运行后的终端输出截图.
提示: 判断测试样例的预测结果时, 可以依照教材实现, 即若有多个分类器预测为正类或者没有分类器预测为正类, 则考虑各分类器的预测置信度 ($f(\boldsymbol{x})$ 之值), 选择置信度最大的类别标记作为分类结果.

(5) **[5pts]** 基于 (4) 绘制并在 `HW2.pdf` 中展示**训练错误率和测试错误率随 λ 变化的折线图**. 注意: 图像横轴为 λ; 训练错误率和测试错误率应当使用不同颜色的曲线.

Solution. 此处用于写解答 (中英文均可)

Acknowledgments

允许与其他同样未完成作业的同学讨论作业的内容, 但需在此注明并加以致谢; 如在作业过程中, 参考了互联网上的资料或大语言模型的生成结果, 且对完成作业有帮助的, 亦需注明并致谢.