

机器学习导论 习题一

学号, 姓名, 邮箱

2023 年 3 月 14 日

作业提交注意事项

1. 请在 LaTeX 模板中第一页填写个人的学号、姓名、邮箱;
2. 本次作业需提交作答后的该 pdf 文件、编程题代码 (.py 文件); **请将二者打包为 .zip 文件上传**. 注意命名规则, 三个文件均命名为“学号 _ 姓名” + “. 后缀” (例如 211300001_ 张三” + “.pdf”、“.py”、“.zip”);
3. 若多次提交作业, 则在命名 .zip 文件时加上版本号, 例如 211300001_ 张三_v1.zip” (批改时以版本号最高的文件为准);
4. 本次作业提交截止时间为 **3 月 29 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; 除特殊情况 (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**;
5. 本次作业提交地址为 [here](#), 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.

1 [15pts] Derivatives of Matrices

有 $\alpha \in \mathbb{R}$, $\mathbf{y} \in \mathbb{R}^{m \times 1}$, $\mathbf{x} \in \mathbb{R}^{n \times 1}$, 试完成下题, 并给出计算过程.

- (1) [4pts] 此问中假设 $\mathbf{A} \in \mathbb{R}^{n \times n}$, 且 $\alpha = \mathbf{x}^\top \mathbf{A} \mathbf{x}$, 试求 $\frac{\partial \alpha}{\partial \mathbf{x}}$.
- (2) [5pts] 此问中假设 $\mathbf{A} \in \mathbb{R}^{m \times n}$, 且 $\alpha = \mathbf{y}^\top \mathbf{A} \mathbf{x}$, 同时 \mathbf{y} 、 \mathbf{x} 为 \mathbf{z} 的函数, 试求 $\frac{\partial \alpha}{\partial \mathbf{z}}$.
- (3) [6pts] 此问中假设 $\mathbf{A} \in \mathbb{R}^{n \times n}$ 且 \mathbf{A} 可逆, \mathbf{A} 为 α 的函数同时 $\frac{\partial \mathbf{A}}{\partial \alpha}$ 已知. 试求 $\frac{\partial \mathbf{A}^{-1}}{\partial \alpha}$.

(提示: 可以参考 The Matrix Cookbook.)

Solution. 此处用于写解答 (中英文均可)

2 [15pts] Performance Measure

性能度量是衡量模型泛化能力的评价标准, 在对比不同模型的能力时, 使用不同的性能度量往往会导致不同的评判结果. 请仔细阅读《机器学习》第二章 2.3.3 节. 在书中, 我们学习并计算了模型的二分类性能度量. 下面我们给出一个多分类 (四分类) 的例子, 请根据学习器的具体表现, 回答如下问题.

表 1: 类别的真实标记与预测

真实类别 \ 预测类别	第一类	第二类	第三类	第四类
第一类	7	2	1	0
第二类	0	9	0	1
第三类	1	0	8	1
第四类	1	2	1	6

- (1) [5pts] 如表 1 所示, 请计算该学习器的错误率及精度.
- (2) [5pts] 请分别计算宏查准率, 宏查全率, 微查准率, 微查全率, 并两两比较大小.
- (3) [5pts] 分别使用宏查准率, 宏查全率, 微查准率, 微查全率计算宏 $F1$ 度量, 微 $F1$ 度量, 并比较大小.

Solution. 此处用于写解答 (中英文均可)

3 [15pts] ROC & AUC

ROC 曲线与其对应的 AUC 值可以反应分类器在“一般情况下”泛化性能的好坏. 请仔细阅读《机器学习》第二章 2.3.3 节, 并完成本题.

表 2: 样例的真实标记与预测

样例	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
标记	0	1	0	1	0	0	1	1	0
分类器输出值	0.4	0.9	0.7	0.4	0.2	0.8	0.8	0.6	0.5

- (1) [5pts] 如表 2 所示, 第二行为样例对应的真实标记, 第三行为某分类器对样例的预测结果. 请根据上述结果, 绘制分类器在该样例集合上的 ROC 曲线, 并写出绘图中使用到的节点 (在坐标系中的) 坐标及其对应的阈值与样例编号.
- (2) [3pts] 根据上题中的 ROC 曲线, 计算其对应的 AUC 值 (请给出具体的计算步骤).
- (3) [7pts] 结合前两问使用的例子 (可以借助图片示意), 试证明对有限样例成立:

$$\text{AUC} = \frac{1}{m^+m^-} \sum_{x^+ \in D^+} \sum_{x^- \in D^-} \left(\mathbb{I}\{f(x^+) > f(x^-)\} + \frac{1}{2} \mathbb{I}\{f(x^+) = f(x^-)\} \right). \quad (3.1)$$

Solution. 此处用于写解答 (中英文均可)

4 [20pts] Linear Regression

线性回归模型是一类常见的机器学习方法，其基础形式与变体常应用在回归任务中。根据《机器学习》第三章 3.2 节中的定义，可以将收集到的 d 维数据及其标签如下表示：

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1d} & 1 \\ x_{21} & x_{22} & \cdots & x_{2d} & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{md} & 1 \end{pmatrix} = \begin{pmatrix} \mathbf{x}_1^\top & 1 \\ \mathbf{x}_2^\top & 1 \\ \vdots & \vdots \\ \mathbf{x}_m^\top & 1 \end{pmatrix}; \quad \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}.$$

将参数项与截距项合在一起，定义为 $\hat{\mathbf{w}} = (\mathbf{w}^\top; b)^\top$ 。此时成立 $\hat{\mathbf{y}} = \mathbf{X}\hat{\mathbf{w}}$ 。《机器学习》式 (3.11) 给出了最小二乘估计 (Least Square Estimator, LSE) 的闭式解：

$$\hat{\mathbf{w}}_{\text{LSE}}^* = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}. \quad (4.1)$$

(1) [8pts] (投影矩阵的性质) 容易验证，当采用最小二乘估计 $\hat{\mathbf{w}}_{\text{LSE}}^*$ 时，成立：

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\mathbf{w}}_{\text{LSE}}^* = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}.$$

记 $\mathbf{H} = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$ ，则有 $\hat{\mathbf{y}} = \mathbf{H}\mathbf{y}$ 。 \mathbf{H} 被称为“Hat Matrix”，其存在可以从空间的角度，把 $\hat{\mathbf{y}}$ 看作是 \mathbf{y} 在矩阵 \mathbf{H} 空间中的投影。 \mathbf{H} 矩阵有着许多良好的性质。已知此时 \mathbf{X} 矩阵列满秩， \mathbf{I} 为单位阵，试求 $\mathbf{I} - \mathbf{H}$ 的全部特征值并注明特征值的重数。

(提示：利用 \mathbf{H} 矩阵的投影性质与对称性。)

(2) [5pts] (岭回归) 当数据量 m 较小或数据维度 d 较高时，矩阵 $\mathbf{X}^\top \mathbf{X}$ 可能不满秩，4.1 中的取逆操作难以实现。此时可使用岭回归代替原始回归问题，其形式如下：

$$\hat{\mathbf{w}}_{\text{Ridge}}^* = \arg \min_{\hat{\mathbf{w}}} \frac{1}{2} (\|\mathbf{y} - \mathbf{X}\hat{\mathbf{w}}\|_2^2 + \lambda \|\hat{\mathbf{w}}\|_2^2). \quad (4.2)$$

试求岭回归问题的闭式解，并简述其对原问题的改进。

(3) [7pts] 定义 $\tilde{\mathbf{x}}_i = (\mathbf{x}_i^\top; 1)^\top$ ， $\hat{y}_i = \tilde{\mathbf{x}}_i^\top \hat{\mathbf{w}}_{\text{LSE}}^*$ ， $\bar{y} = \frac{1}{m} \sum_{i=1}^m y_i$ 。

对线性回归模型进行统计分析时，会涉及如下三个基础定义：

$$\begin{cases} \text{Total sum of squares (SST):} & \sum_{i=1}^m (y_i - \bar{y})^2 \\ \text{Regression sum of squares (SSR):} & \sum_{i=1}^m (\hat{y}_i - y_i)^2 \\ \text{Residual sum of squares (SSE):} & \sum_{i=1}^m (\hat{y}_i - \bar{y})^2 \end{cases}$$

试证明 $\text{SST} = \text{SSR} + \text{SSE}$ 。(提示：使用向量形式可以简化证明步骤。)

Solution. 此处用于写解答 (中英文均可)

5 [35pts] Logistic Regression in Practice

对数几率回归 (Logistic Regression, 简称 LR) 是实际应用中非常常用的分类学习算法.

- (1) [30pts] 请编程实现二分类的 LR, 要求采用牛顿法进行优化求解. 详细编程题指南请参见链接: [here](#). 请将绘制好的 ROC 曲线放在解答处, 并记录模型的精度与 AUC (保留 4 位小数).
- (2) [5pts] 试简述在对数几率回归中, 相比梯度下降方法, 使用牛顿法的优点和缺点.

Solution. 此处用于写解答 (中英文均可)