机器学习入门，2023 年秋季学期作业 4

(截止时间：12月19日（星期二）晚上11:59（美国中部标准时间）

1. [15 分] [最大边际分类器] 考虑由 nd 维样本点组成的数据集 $\{X_1, \ldots, X_n\}$。每个样本点 $X_i \in \mathbb{R}^d$ 都有一个相应的标签 y_i，表示该点属于哪个类别。现在，我们假设只有两个类别，每个点要么属于给定类别（$y_i = 1$），要么不属于该类别（$y_i =-1$）。考虑由超平面定义的线性决策边界

$$H = x \in \mathbb{R}^d : x \cdot w + \alpha = 0\}。$$

最大边际分类器最大化线性决策边界到边界两侧最近训练点的距离，同时正确分类所有训练点。

(a) 如果类内样本点位于决策边界的正边，则该样本点被正确分类；如果类外样本点位于决策边界的负边，则该样本点被正确分类。请写出一组 n 个约束条件，以确保所有 n 个点都被正确分类。[3 分]

(b) 最大边际分类器的目标是最大化训练点到决策边界的距离。推导点 X_i 到超平面 H 的距离 [3 分]。

(c) 假设所有点都被正确分类，请写出一个不等式，仅用法线向量 w 来表示样本点 X_i 与超平面 H 的距离。[3 分]。

(d) 对于最大边际分类器来说，边界两侧最接近决策边界的训练点被称为支持向量。从任何支持向量到决策边界的 距离是多少？[3 分]

(e) 利用前面的部分，写出最大边际分类器的优化问题。[3 分]

解决方案

2. [15 分] 考虑由 n 个观测值 $X \in R^{n\times d}$ 组成的数据集，我们的目标是将数据投影到维度为 p 的子空间，$p < d$。证明基于投影方差最大化的 PCA 等价于基于投影误差（欧氏误差）最小化的 PCA。

解决方案

3. [15 分] [手工进行 PCA 分析] 让我们来进行主成分分析（PCA）！ 考虑由六个点组成的样本 $X_i \in \mathrm{R}^2$ 。

0 0 1 1 , 2 2

0 1 ’ 0 ’ ’ 2 ’ ˙1 2

(a) [4 分] 计算样本点的平均值，并写出居中设计矩阵 $\dot{X}$。提示：样本平均值为

提示：从每个样本中减去平均值，就形成了居中设计矩阵

$$\dot{X} =$$

(b) [5 分] 找出该样本的所有主成分。将它们写成单位向量。提示：我们数据集的主

成分是矩阵的特征向量

$$\dot{X}^{\top}\dot{X} =$$

该对称矩阵的特征多项式为

$$\det sI - X^{\top} X$$

(c) [6分]

如果只使用一个主成分，您会首选这两个主成分中的哪一个？[2 分]

PCA 算法使用哪些信息来决定一个主成分优于另一个主成分？[2 分]

从优化的角度看，我们为什么更喜欢那一种？[2 分]

解决方案

4. [15 分] [算术表达式的反向传播]考虑一个算术网络，其输入为

a、b 和 c，计算以下操作序列，其中 $s(\gamma) = \frac{1}{1+e^{-\gamma}}$ 是对数（sigmoid）函数，$r(\gamma) = \max\{0, \gamma\}$ 是 ReLU 使用的铰链函数。

$$d = ab \quad e = s(d) \quad f = r(a) \quad g = 3a \quad h = 2e + f + g \quad i = ch \quad j = f + i^2$$

我们想用反推法求得 j 相对于其他变量 a 至 i 的偏导数。这意味着，对于每个变量 z，我们希望你用两种形式写出 $\partial j/\partial z$：(1) 涉及直接使用 z 值的每个变量的导数，(2) 输入和中间值 $a \dots i$，*尽可能*简单，但不使用导数符号。例如，我们可以写成

$$\frac{\partial j}{\partial i} = \frac{dj}{di} = 2i \quad \text{(仅此一项无需链式规则)}$$

$$\frac{\partial j}{\partial h} = \frac{\partial j}{\partial i}\frac{\partial i}{\partial h} = 2ic \quad \text{(链式法则，然后反推导数表达式)}$$

(a) 现在，请写出 $\partial j/\partial g$、$\partial j/\partial f$、$\partial j/\partial e$、$\partial j/\partial d$、$\partial j/\partial c$、$\partial j/\partial b$ 和 $\partial j/\partial a$ *的*表达式，就像我们写的那样上面的 $\partial j/\partial h$。如果需要，用 $s(\gamma)$表示导数 $s'(\gamma)$，用指示函数 $1(\gamma \geq 0)$表示导数 $r'(\gamma)$。(提示：两处使用了 f，三处使用了 a，因此需要使用多元链式法则。将网络绘制成有向图可能会有帮助，但不是必需的）。

解决方案