

人工智能导论第二次作业

2022年4月

1 第一题(10分)

判断下列说法是否正确并简述理由:

- (1) 相较于L2正则化, 使用L1正则化会得到更为稀疏的参数向量。(2分)
- (2) 交叉验证过程中需要使用训练数据和测试数据。(2分)
- (3) 在K近邻方法中, 当K减小时, 模型的分类面会更加光滑, 不易发生过拟合。(2分)
- (4) 决策树算法只可应用于分类问题。(2分)
- (5) 随机森林中, 自举 (Bootstrap) 有助于降低基学习器的偏差。(2分)

2 第二题(15分)

考虑带权重的支持向量机, 其中 $C > 0$, 权重常数 $\gamma_i \in (0, +\infty)$:

$$\min_{\mathbf{w}, b, \xi} \frac{1}{2} \|\mathbf{w}\|_2^2 + C \sum_{i=1}^n \gamma_i \xi_i$$

$$\text{s.t.} \quad y_i (\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 - \xi_i, \xi_i \geq 0, 1 \leq i \leq n.$$

请使用次梯度下降法 (Subgradient descent) 求解上述最小化问题, 并给出完整计算公式和伪代码。(提示: 可把问题形式转换为Hinge loss与L2正则项的形式)

3 第三题(15分)

考虑以下一个多层感知机, 其中输入维度为 n , 隐藏层维度为 m , 输出为 k 维的概率分布, 其前向传播如下:

$$\mathbf{z}_1 = \mathbf{W}^{(1)} \mathbf{x} + \mathbf{b}^{(1)}$$

$$\mathbf{h}_1 = \text{ReLU}(\mathbf{z}_1)$$

$$\mathbf{z}_2 = \mathbf{W}^{(2)} \mathbf{h}_1 + \mathbf{b}^{(2)}$$

$$\hat{\mathbf{y}} = \text{Softmax}(\mathbf{z}_2)$$

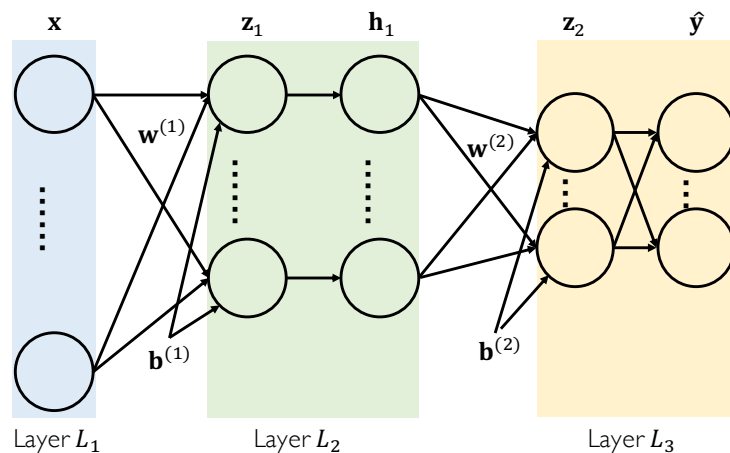


Figure 1: 多层感知机的网络结构

其中 $\mathbf{x} \in \mathbb{R}^{n \times 1}$, $\mathbf{W}^{(1)} \in \mathbb{R}^{m \times n}$, $\mathbf{z}_1, \mathbf{b}^{(1)}, \mathbf{h}_1 \in \mathbb{R}^{m \times 1}$, $\mathbf{W}^{(2)} \in \mathbb{R}^{k \times m}$, $\mathbf{z}_2, \mathbf{b}^{(2)}, \hat{\mathbf{y}} \in \mathbb{R}^{k \times 1}$ 。

前向传播之后，我们使用交叉熵损失函数：

$$f_{\text{CE}}(\mathbf{W}^{(1)}, \mathbf{b}^{(1)}, \mathbf{W}^{(2)}, \mathbf{b}^{(2)}) = - \sum_{i=1}^k \mathbf{y}_i \log \hat{\mathbf{y}}_i$$

其中 \mathbf{y} 是一个只有一个元素为1，其他元素均为0的 $k \times 1$ 维One-Hot向量。

试求其反向传播过程中 $\frac{\partial f_{\text{CE}}}{\partial \mathbf{W}^{(2)}}$, $\frac{\partial f_{\text{CE}}}{\partial \mathbf{b}^{(2)}}$, $\frac{\partial f_{\text{CE}}}{\partial \mathbf{W}^{(1)}}$, $\frac{\partial f_{\text{CE}}}{\partial \mathbf{b}^{(1)}}$ 的值。(提示：参考矩阵求导术)

4 第四题(60分)

编程题: 相关材料见./classification文件夹，作业详细要求见./classification/README.md。