人工智能导论第二次作业问答题

祝尔乐 2020013020

2022 年 4 月 28 日

1　第一题

(1)

正确。
理由：写出加入正则化的求解式的等价形式：

$$\hat{\mathbf{w}} = \underset{||\mathbf{w}||_k \leq r}{\arg\min} \sum_{i=1}^{n} (\mathbf{w}^T \mathbf{x}_i - y_i)^2, k = 1, 2 \tag{1}$$

上式求出的解 w^* 可以看做由 Lk 范数 (k=1,2) 规定的半径为 r 的球面与曲面 $\sum_{i=1}^{n} (\mathbf{w}^T \mathbf{x}_i - y_i)^2 = Const$ 的交点，相对于 L2 范数，L1 范数与曲面的交点落在顶点的概率较大，即交点更容易落在坐标轴上，于是得到的 w^* 较为稀疏。

(2)

错误。
交叉验证中将用于训练的数据分为 k 折，然后循环的选取训练集和验证集进行训练和验证，而测试数据是不参与训练的，于是交叉验证中不需要使用测试数据。

(3)

错误。
K 邻近方法中 K 越小，模型的分类面越容易产生由噪声引起的畸变，变得不光滑。

(4)

错误。
决策树也可以用于回归问题，例如最小二乘决策树。

(5)

错误。
随机森林中，自举是相对于多次取样而言的，对基学习器而言，仍然是单次采样，偏差不受自举影响，自举可以降低集成学习器的偏差。

2 第二题

消去 ξ, 并令 $\mathbf{W} \leftarrow [\mathbf{w}, b], \mathbf{X} = [\mathbf{x}, 1]$, 将问题形式转化为：

$$\min_{\mathbf{W}} f(\mathbf{W}) \ \textit{while} \ f(\mathbf{W}) = \frac{1}{2}||\mathbf{W}||_2^2 + \frac{C}{n}\sum_{i=1}^{n}\gamma_i \max\{0, 1 - y_i(\mathbf{W}^T\mathbf{X}_i)\} \tag{2}$$

对上式求解次梯度：

$$\bigtriangledown f = \frac{1}{2}\mathbf{W} - \frac{C}{n}\sum_{i=1}^{n}\gamma_i I[y_i(\mathbf{W}^T\mathbf{X}_i) < 1]y_i\mathbf{X}_i \tag{3}$$

使用梯度下降更新, 设 $\mathbf{W}_t, Batch(t)$ 为 t 时刻的权值和训练样本的下标集，则：

$$\begin{aligned} \mathbf{W}_{t+1} &= \mathbf{W}_t - \eta_t \bigtriangledown_t f & (4) \\ &= (1 - \frac{1}{2}\eta_t)\mathbf{W}_t + \eta_t\frac{C}{n}\sum_{i\in Batch(t)}\gamma_i I[y_i(\mathbf{W}_t^T\mathbf{X}_i) < 1]y_i\mathbf{X}_i & (5) \end{aligned}$$

写出算法伪代码：

Algorithm 1: `SubGD-WeightSVM`(S, T, k)

Input: S(train set), T(max iteration num), k(batch num)

Output: a approximation of $\mathbf{W}$

Initialize $\mathbf{W}$ with $\mathbf{0}$

for $t \leftarrow 1$ **to** T **do**

- Choose a random batch of size k from S as A(index)
- $A^+ \leftarrow \{i \in A : y_i(\mathbf{W}_i^T\mathbf{X}_i) < 1\}$
- $\eta \leftarrow \frac{2}{t}$
- $\mathbf{W} \leftarrow (1 - \frac{1}{2}\eta)\mathbf{W} - \eta\frac{C}{k}\sum_{i\in A^+}\gamma_i y_i \mathbf{X}_i$

end

return $\boldsymbol{W}$

3 第三题

推导如下：

3\. 设 $l = f_{CE}\left(W^{(1)}, b^{(1)}, W^{(2)}, b^{(2)}\right) = -y^T \log \mathrm{Softmax}\, z_2$

由 $\mathrm{Softmax}(x) = \dfrac{\exp(x)}{\mathbf{1}^T \exp(x)}$

$$\Rightarrow l = -y^T \left[\log \exp(z_2) - \mathbf{1} \log\left(\mathbf{1}^T \exp(z_2)\right)\right]$$

$$= -y^T z_2 + \log\left(\mathbf{1}^T \exp(z_2)\right) \quad (y^T \mathbf{1} = 1,\ y \text{ 为 one-hot vector})$$

l 对 z_2 求导：$dl = -y^T dz_2 + \dfrac{\exp(z_2)^T dz_2}{\mathbf{1}^T \exp(z_2)} = [\mathrm{Softmax}(z_2) - y]^T dz_2 \Rightarrow \dfrac{\partial l}{\partial z_2} = \mathrm{Softmax}(z_2) - y$

由 $z_2 = W^{(2)} h_1 + b^{(2)}$

$$\Rightarrow dl = tr\left(\frac{\partial l}{\partial z_2}^T dz_2\right) = tr\left(\frac{\partial l}{\partial z_2}^T dW^{(2)} h_1\right) + tr\left(\frac{\partial l}{\partial z_2}^T W^{(2)} dh_1\right) + tr\left(\frac{\partial l}{\partial z_2}^T db^{(2)}\right)$$

可交换

$$\Rightarrow \begin{cases} \dfrac{\partial l}{\partial W^{(2)}} = \dfrac{\partial l}{\partial z_2} h_1^T = [\mathrm{Softmax}(z_2) - y]\, h_1^T & \left(\text{即 } \dfrac{\partial f_{CE}}{\partial W^{(2)}}\right) \\ \dfrac{\partial l}{\partial b^{(2)}} = \dfrac{\partial l}{\partial z_2} = \mathrm{Softmax}(z_2) - y & \left(\text{即 } \dfrac{\partial f_{CE}}{\partial b^{(2)}}\right) \end{cases}$$

下求 $\dfrac{\partial l}{\partial W^{(1)}}$，$\dfrac{\partial l}{\partial b^{(1)}}$，此过程可令 $dW^{(2)} = 0$，$db^{(2)} = 0$

$$dl = tr\left(\frac{\partial l}{\partial z_2}^T W^{(2)} dh_1\right) = tr\left(\frac{\partial l}{\partial h_1}^T dh_1\right)$$

$$\Rightarrow \frac{\partial l}{\partial h_1} = W^{(2)T} \frac{\partial l}{\partial z_2}$$

由 $h_1 = \mathrm{ReLU}(z_1)$

$$dl = tr\left(\frac{\partial l}{\partial h_1}^T \left(\mathrm{ReLU}'(z_1) \odot dz_1\right)\right) = tr\left(\left(\frac{\partial l}{\partial h_1} \odot \mathrm{ReLU}'(z_1)\right)^T dz_1\right)$$

$$\Rightarrow \frac{\partial l}{\partial z_1} = \frac{\partial l}{\partial h_1} \odot \mathrm{ReLU}'(z_1)$$

由 $z_1 = W^{(1)} x + b^{(1)}$

$$\Rightarrow dl = tr\left(\frac{\partial l}{\partial z_1}^T dz_1\right) = tr\left(\frac{\partial l}{\partial z_1}^T dW^{(1)} x\right) + tr\left(\frac{\partial l}{\partial z_1}^T W^{(1)} dx\right) + tr\left(\frac{\partial l}{\partial z_1}^T db^{(1)}\right)$$

可交换

$$\Rightarrow \frac{\partial l}{\partial W^{(1)}} = \frac{\partial l}{\partial z_1} x^T = \left[\frac{\partial l}{\partial h_1} \odot \mathrm{ReLU}'(z_1)\right] x^T = \left[\left(W^{(2)T}(\mathrm{Softmax}(z_2) - y)\right) \odot \mathrm{ReLU}'(z_1)\right] x^T \quad \left(\text{即 } \frac{\partial f_{CE}}{\partial W^{(1)}}\right)$$

$$\frac{\partial l}{\partial b^{(1)}} = \frac{\partial l}{\partial z_1} = \left(W^{(2)T}(\mathrm{Softmax}(z_2) - y)\right) \odot \mathrm{ReLU}'(z_1) \quad \left(\text{即 } \frac{\partial f_{CE}}{\partial b^{(1)}}\right)$$

式中 $\mathrm{ReLU}'(x) = \begin{bmatrix} \mathbb{I}[x_1 > 0] \\ \vdots \\ \mathbb{I}[x_i > 0] \\ \vdots \\ \mathbb{I}[x_n > 0] \end{bmatrix}$

4 第四题

在 classification 文件夹中。