扬州大学作业纸

班级 计科2002 姓名 李蔚 学号 202801108 日期 第 1 页

第三章作业：

1. B
2. C
3. D
4. B
5. ACD
6. ABCD
7. ACD

8. 简述线性分类器训练的一般思路是什么？

解：(1) 数据预处理，收集和准备用于训练和测试的数据集。

(2) 选择合适的线性分类模型，并定义模型的输入和输出。

(3) 定义损失函数：损失函数用于衡量模型在训练过程中的性能，即模型预测的结果与真实结果之间的差距。

(4) 选择优化算法，线性分类器常用的算法有梯度下降法等

(5) 训练模型：在训练过程中，需要将训练集的数据点输入到模型中进行预测，并计算损失函数的值，使用梯度下降法等优化算法对模型参数进行更新。

(6) 模型评估：训练完成后，需要使用测试集的数据点对模型进行评估

9. 简述广义线性化的思想。

解：广义线性化的思想是将非线性问题转化为线性问题，以方便使用线性模型来解决。具体来说，广义线性化方法将非线性函数通过泰勒展开的方式近似为一阶线性函数，从而使问题变为线性问题。从另一个方面来说，广义线性化可以方便地将线性模型推广到更广泛的情况，包括非线性和离散型数据的建模，从而可以提高模型的预测和泛化能力，是一种很实用的模型。

扬州大学作业纸

班级 计科2002　姓名 李蔚　学号 201801108　日期　第 2 页

10. 给定一个训练集，其中正例样本为 $x_1=(3,3)$，$x_2=(5,2)$，负例样本为 $x_1=(1,1)$，请用感知机算法求出能正确分类该数据集的分类模型：$f(x)=\mathrm{sign}(W^Tx+b)$，其中 $W=(W^{(1)},W^{(2)})^T$，$x=(x^{(1)},x^{(2)})^T$

请使用随机梯度下降法更新，W初始化为零向量，取学习步长 $\eta=1$，请写出详细的计算过程。

解：(1) 对训练数据进行规范化得到

$x_1=(3,3,1)$，$y_1=+1$

$x_2=(5,2,1)$，$y_2=+1$

$x_3=(1,1,1)$，$y_3=-1$

(2) 利用随机梯度下降法求解

初始化 $W=0$，取学习步长 $\eta=1$

对于 x_1，$y_1W^Tx_1=0$，修正权向量 $W=W+\eta y_1x_1=(3,3,1)$

对于 x_2，$y_2W^Tx_2>0$，无需修正

对于 x_3，$y_3W^Tx_3<0$，修正权向量 $W=W+\eta y_3x_3=(2,2,0)$

对于 x_1，$y_1W^Tx_1>0$，无需修正

对于 x_2，$y_2W^Tx_2>0$，无需修正

对于 x_2，$y_3W^Tx_3<0$，修正权向量 $W=W+\eta y_3x_3=(1,1,-1)$

对于 x_1，$y_1W^Tx_1>0$，无需修正

对于 x_2，$y_2W^Tx_2>0$，无需修正

对于 x_3，$y_3W^Tx_3<0$，修正权向量 $W=W+\eta y_3x_3=(0,0,-2)$

对于 x_1，$y_1W^Tx_1<0$，修正权向量 $W=W+\eta y_1x_1=(3,3,-1)$

对于 x_2，$y_2W^Tx_2>0$，无需修正

对于 x_3，$y_3W^Tx_3<0$，修正权向量 $W=W+\eta y_3x_3=(2,2,-2)$

对于 x_1，$y_1W^Tx_1>0$，无需修正

对于 x_2，$y_2W^Tx_2>0$，无需修正

对于 x_3，$y_3W^Tx_3<0$，修正权向量 $W=W+\eta y_3x_3=(1,1,-3)$

对于 x_1，$y_1W^Tx_1>0$，无需修正

扬州大学作业纸

班级 计科2002 姓名 李清 学号 20280118 日期 第 3 页

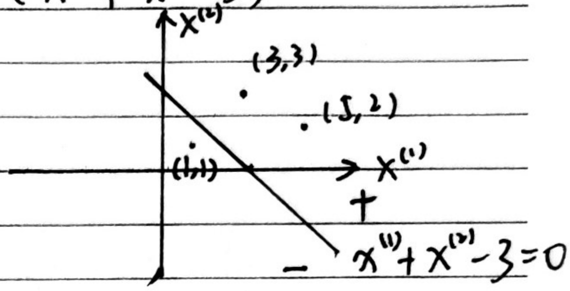
对于 X_2，$y_2 W^T X_2 > 0$，无需修正

对于 X_3，$y_3 W^T X_3 > 0$，无需修正

算法终止

得到的感知机模型为 $f(x) = \mathrm{sign}(x^{(1)} + x^{(2)} - 3)$

其中 $(x^{(1)}, x^{(2)})$ 为未扩充的实例



扬州大学作业纸

班级 计科2002　姓名 李蔚　学号 202801108　日期　　第 1 页

第4章作业：

1. D
2. A
3. C
4. C
5. ABCD
6. ABCD
7. ABD
8. 线性支持向量机的优化问题还可以表示为如下形式：

$$\min_{w,b,\varepsilon} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\varepsilon_i^2$$

$$s.t.\ y_i(w^T x_i + b) \geq 1-\varepsilon_i,\ i=1\cdots N$$

(1) 请推导出该优化问题的对偶问题。

(2) 假设对偶问题最优解为 $\alpha^* = (\alpha_1^* \cdots \alpha_N^*)$，请利用KKT条件构造出原始问题的最优解 w^*, b^*。

解：(1) 定义拉格朗日函数为：

$$L(w,b,\varepsilon,\alpha,\beta) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\varepsilon_i + \sum_{i=1}^{N}\alpha_i(1-\varepsilon_i - y_i(w^T x_i + b)) - \sum_{i=1}^{N}\beta_i\varepsilon_i$$

对偶优化目标函数为：

$$\theta_D(\alpha,\beta) = \min_{w,b,\varepsilon} L(w,b,\varepsilon,\alpha,\beta)$$

令 $L(w,b,\varepsilon,\alpha,\beta)$ 关于 w, b, ε_i 的偏导数分别为 0：

$$\partial L/\partial w = 0 \Rightarrow w = \sum_{i=1}^{N}\alpha_i y_i x_i$$

$$\partial L/\partial b = 0 \Rightarrow \sum_{i=1}^{N}\alpha_i y_i = 0$$

$$\partial L/\partial \varepsilon_i = 0 \Rightarrow C = \alpha_i + \beta_i \quad i=1,2\cdots N$$

回代，可以得到

$$\theta_D(\alpha,\beta) = \frac{1}{2}\left(\sum_{i=1}^{N}\alpha_i y_i x_i\right)^T\left(\sum_{i=1}^{N}\alpha_i y_i x_i\right) + \cancel{C\sum_{i=1}^{N}\varepsilon_i} + \sum_{i=1}^{N}\alpha_i - \cancel{\sum_{i=1}^{N}\alpha_i\varepsilon_i}$$

$$- \left(\sum_{i=1}^{N}\alpha_i y_i x_i\right)^T\left(\sum_{i=1}^{N}\alpha_i y_i x_i\right) - \cancel{b\sum_{i=1}^{N}\alpha_i y_i} - \cancel{\sum_{i=1}^{N}\beta_i\varepsilon_i}$$

$$= -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j x_i^T x_j + \sum_{i=1}^{N}\alpha_i$$

扬州大学作业纸

班级 计科2002 姓名 李楠 学号 203801108 日期 第 2 页

$\because \arg\max_{\alpha,\beta:\alpha\geq 0,\beta\geq 0} \theta_D(\alpha,\beta) \Leftrightarrow \arg\min_{\alpha,\beta:\alpha\geq 0,\beta\geq 0} -\theta_D(\alpha,\beta)$

对偶优化问题为：

$$\min_{\alpha,\beta} \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j x_i^T x_j - \sum_{i=1}^{N}\alpha_i$$

$$\text{s.t.} \sum_{i=1}^{N}\alpha_i y_i = 0$$

$$\alpha_i + \beta_i = C \quad i=1,2\cdots N$$

$$\alpha_i \geq 0, \beta_i \geq 0, \quad i=1,2\cdots N$$

简化后，得到原问题的对偶优化问题为

$$\min_{\alpha} \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j x_i^T x_j - \sum_{i=1}^{N}\alpha_i$$

$$\text{s.t.} \sum_{i=1}^{N}\alpha_i y_i = 0$$

$$0 \leq \alpha_i \leq C, \quad i=1,2\cdots N$$

(2) 由KKT条件可得

$$\partial L/\partial w = 0 \quad \Rightarrow \quad w^* = \sum_{i=1}^{N}\alpha_i^* y_i x_i$$

$$\partial L/\partial b = 0 \quad \Rightarrow \quad \sum_{i=1}^{N}\alpha_i^* y_i = 0$$

$$\partial L/\partial \varepsilon_i = 0 \quad \Rightarrow \quad \alpha_i^* + \beta_i^* = C \quad i=1,2\cdots N$$

对偶互补条件：$\alpha_i^*(1-\varepsilon_i^* - y_i(w^{*T}x_i + b^*)) = 0 \quad i=1,2\cdots N$

$\beta_i^*\varepsilon_i^* = 0 \quad i=1,2\cdots N$

原问题约束：$y_i(w^{*T}x_i + b^*) \geq 1-\varepsilon_i^* \quad i=1,2\cdots N$

$\varepsilon_i \geq 0 \quad i=1,2\cdots N$

对偶约束：$\sum_{i=1}^{N}\alpha_i^* y_i = 0$

$0 \leq \alpha_i^* \leq C \quad i=1,2\cdots N$

由对偶互补条件：$\left.\begin{array}{l}\text{若 } 0<\alpha_j^*<C, \text{则 } y_j(w^{*T}x_j+b) = 1-\varepsilon_j^* \\ \text{若 } 0<\alpha_j^*<C, \text{则 } \beta_j^*>0 \Rightarrow \varepsilon_j = 0\end{array}\right\} \Rightarrow y_j(w^{*T}x_j + b^*) = 1$

结合 $y_j^2 = 1$ 得 $b^* = y_j - w^{*T}x_j = y_j - \sum_{i=1}^{N}\alpha_i^* y_i x_i^T x_j$

综上，$w^* = \sum_{i=1}^{N}\alpha_i^* y_i x_i$

$b^* = y_j - \sum_{i=1}^{N}\alpha_i^* y_i x_i^T x_j$