2_广义线性模型

2_3_感知机模型

2_3_1_感知机模型与分类超平面

假设输入空间$\mathcal{X} \subseteq \mathbb{R}^n$，输出空间$\mathcal{Y} = \{+1, -1\}$。输入$\mathbf{x} \in \mathcal{X}$表示实例的特征向量，对应于输入空间的点；输出$y \in \mathcal{Y}$表示实例的类别。由输入空间到输出空间的函数

$$f(\mathbf{x}) = sign(\mathbf{w} \cdot \mathbf{x} + b)$$

称为感知机。其中，$\mathbf{w}$和b为感知机模型参数，$\mathbf{w} \in \mathbb{R}^n$叫做权值或权值向量，$b \in R$叫偏置，$\mathbf{w} \cdot \mathbf{x}$表示$\mathbf{w}$和$\mathbf{x}$的内积。$sign$是符号函数，即

$$sign(x) = \begin{cases} +1, x \geq 0 \\ -1, x < 0 \end{cases}$$

感知机是一种线性分类模型。感知机模型的假设空间是定义在特征空间中的所有线性分类模型或线性分类器，即函数集合$\{f | f(\mathbf{x}) = \mathbf{w} \cdot x + b\}$。

给定数据集

$$D = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \cdots, (\mathbf{x}_N, y_N)\}$$

其中，$\mathbf{x}_i \in \mathcal{X} = \mathbb{R}^n, y_i \in \mathcal{Y} = \{+1, -1\}, i = 1, 2, \cdots, N$，如果存在某个超平面$S$

$$\mathbf{w} \cdot \mathbf{x} + b = 0$$

能够将数据集的正实例和负实例完全正确地划分到超平面的两侧，即对所有$y_i = +1$的实例$\mathbf{x}_i$，有$\mathbf{w} \cdot \mathbf{x}_i + b > 0$，对所有$y_i = -1$的实例$\mathbf{x}_i$，有$\mathbf{w} \cdot \mathbf{x}_i + b < 0$，则称数据集$D$为线性可分数据集；否则，称数据集$D$线性不可分。

线性方程

$$\mathbf{w} \cdot \mathbf{x} + b = 0$$

对应于特征空间$\mathcal{X}$中的一个超平面S，其中$\mathbf{w}$是超平面的法向量，b是超平面的截距。超平面S将特征空间划分为两部分，位于其中的点被分为正、负两类，超平面S称为分离超平面。

2_3_2_感知机模型参数学习

输入空间$\mathcal{X}$中的任一点x_0到超平面S的距离：

$$\frac{1}{\|\mathbf{w}\|} |\mathbf{w} \cdot x_0 + b|$$

其中$\|\mathbf{w}\|$是权值向量$\mathbf{w}$的L_2范数。

对于误分类数据$(\mathbf{x}_i, y_i)$，当$\mathbf{w} \cdot \mathbf{x} + b > 0$时，$y_i = -1$，$(\mathbf{x}_i, y_i)$，当$\mathbf{w} \cdot \mathbf{x} + b < 0$时，$y_i = +1$，则有

$$-y_i \left(\mathbf{w} \cdot \mathbf{x}_i + b \right) > 0$$

误分类点$\mathbf{x}_i$到分离超平面的距离:

$$-\frac{1}{\|\mathbf{w}\|} y_i \left(\mathbf{w} \cdot \mathbf{x}_i + b \right)$$

假设超平面S的误分类点集合为M，则所有误分类点到超平面S的总距离：

$$-\frac{1}{\|\mathbf{w}\|} \sum_{\mathbf{x}_i \in M} y_i \left(\mathbf{w} \cdot \mathbf{x}_i + b \right)$$

给定训练数据集

$$D = \left\{ \left(\mathbf{x}_1, y_1 \right), \left(\mathbf{x}_2, y_2 \right), \cdots, \left(\mathbf{x}_N, y_N \right) \right\}$$

其中，$\mathbf{x}_i \in \mathcal{X} = R^n, y_i \in \mathcal{Y} = \left\{ +1, -1 \right\}, i = 1, 2, \cdots, N$。感知机$sign \left(\mathbf{w} \cdot \mathbf{x} + b \right)$的损失函数定义为

$$L \left(\mathbf{w}, b \right) = -\sum_{\mathbf{x}_i \in M} y_i \left(\mathbf{w} \cdot \mathbf{x}_i + b \right)$$

其中，M为误分类点的集合。

给定训练数据集

$$D = \left\{ \left(\mathbf{x}_1, y_1 \right), \left(\mathbf{x}_2, y_2 \right), \cdots, \left(\mathbf{x}_N, y_N \right) \right\}$$

其中，$\mathbf{x}_i \in \mathcal{X} = \mathbb{R}^n, y_i \in \mathcal{Y} = \left\{ +1, -1 \right\}, i = 1, 2, \cdots, N$。求最优参数$\mathbf{w}^*$和$b^*$，使其为以下损失函数极小化问题的解

$$\min_{\mathbf{w}, b} L \left(\mathbf{w}, b \right) = -\sum_{\mathbf{x}_i \in M} y_i \left(\mathbf{w} \cdot \mathbf{x}_i + b \right)$$

其中，M为误分类点的集合。

假设误分类点集合M是固定的，则损失函数$L \left(\mathbf{w}, b \right)$的梯度

$$\begin{aligned} \nabla_{\mathbf{w}} L \left(\mathbf{w}, b \right) &= -\sum_{\mathbf{x}_i \in M} y_i \mathbf{x}_i \\ \nabla_b L \left(\mathbf{w}, b \right) &= -\sum_{\mathbf{x}_i \in M} y_i \end{aligned}$$

随机选取一个误分类点$(\mathbf{x}_i, y_i)$，对$\mathbf{w}, b$进行更新：

$$\begin{aligned} \mathbf{w} &\leftarrow \mathbf{w} + \eta y_i \mathbf{x}_i \\ b &\leftarrow b + \eta y_i \end{aligned}$$

其中，$\eta \left(0 < \eta \leq 1 \right)$是步长，称为学习率。

2_3_3_感知机算法

感知机算法（原始形式）：
输入：训练数据集$D=\{(\mathbf{x}_1,y_1),(\mathbf{x}_2,y_2),\cdots,(\mathbf{x}_N,y_N)\}$，其中$\mathbf{x}_i\in\mathcal{X}=\mathbb{R}^n, y_i\in\mathcal{Y}=\{+1,-1\}, i=1,2,\cdots,N$；学习率$\eta\,(0<\eta\leq 1)$。
输出：$\mathbf{w},b$；感知机模型$f(\mathbf{x})=sign(\mathbf{w}\cdot\mathbf{x}+b)$

1. 选取初值$\mathbf{w}_0,b_0$
2. 在训练集中选取数据$(\mathbf{x}_i,y_i)$
3. 如果$y_i(\mathbf{w}\cdot\mathbf{x}_i+b)\leq 0$

$$\mathbf{w}\leftarrow\mathbf{w}+\eta y_i x_i$$
$$b\leftarrow b+\eta y_i$$

4. 转至2，直至训练集中没有误分类点。

设$\mathbf{w},b$修改n次，则$\mathbf{w},b$关于$(\mathbf{x}_i,y_i)$的增量分别是$\alpha_i y_i\mathbf{x}_i$和$\alpha_i y_i$，其中$\alpha_i=n_i\eta$。$\mathbf{w},b$可表示为

$$\mathbf{w}=\sum_{i=1}^N\alpha_i y_i\mathbf{x}_i$$
$$b=\sum_{i=1}^N\alpha_i y_i$$

其中，$\alpha_i\geq 0, i=1,2,\cdots,N$

感知机算法（对偶形式）：
输入：训练数据集$D=\{(\mathbf{x}_1,y_1),(\mathbf{x}_2,y_2),\cdots,(\mathbf{x}_N,y_N)\}$，其中$\mathbf{x}_i\in\mathcal{X}=R^n, y_i\in\mathcal{Y}=\{+1,-1\}, i=1,2,\cdots,N$；学习率$\eta\,(0<\eta\leq 1)$。
输出：α,b；感知机模型$f(x)=sign\left(\sum_{j=1}^N\alpha_j y_j x_j\cdot x+b\right)$，其中$\alpha=(\alpha_1,\alpha_2,\cdots,\alpha_N)^T$

1. $\alpha\leftarrow 0, b\leftarrow 0$
2. 在训练集中选取数据(x_i,y_i)
3. 如果$y_i\left(\sum_{j=1}^N\alpha_j y_j x_j\cdot x_i+b\right)\leq 0$

$$\alpha_i\leftarrow\alpha_i+\eta$$
$$b\leftarrow b+\eta y_i$$

4. 转至2，直至训练集中没有误分类点。