

数据挖掘-多层感知机

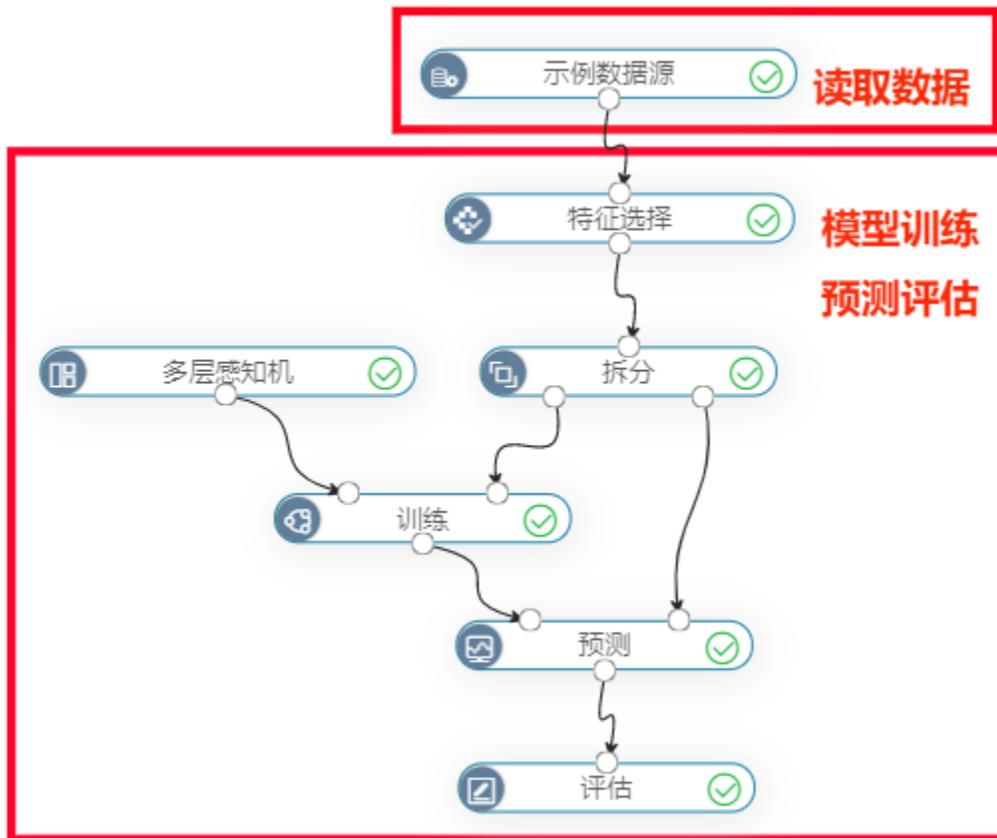
概述

一种基于前馈人工神经网络的分类器。多层感知机由多层节点组成，每个层都完全连接到网络中的下一层，输入层中的节点表示输入数据。所有其他节点通过输入与节点权重 w 和偏移 b 的线性组合将输入映射到输出，并使用激活函数。

- 概述
- 示例
- 参数设置
 - 自动调参设置
 - 示例

示例

使用“动物种类划分”案例数据，预测动物的种类。



参数设置

多层感知机的参数如下：

参数名称	值	说明	
归一化	方法选择	正则化	详情请参考 归一化 介绍说明。
		标准化	
		最小最大值归一化	
		最大绝对值归一化	
参数	单位标准差归一化	勾选后，归一化后数据的标准差等于1	
	平均数据中心化	勾选后，归一化后数据的均值等于0	
自动调参设置	系统默认的各项参数值范围。	必须结合“启用自动调参”功能使用。系统将对设置指定或范围内的参数值循环调参，匹配出最优的组合。详情请参考 自动调参设置 。	
启用自动调参	勾选该项，则系统自动调参数，不需要用户手工设置参数。		
最大迭代数	参数范围为：大于等于10且小于500的整数，默认值为10	算法的最大迭代次数，达到最大迭代次数即退出。 最大迭代次数的值越大，模型训练更充分，但会耗费更多时间。	

感知机的隐藏层节点数量	每层神经元的数量	多层感知机隐藏层的层数及对应每一层的神经元的个数。 以英文逗号分隔，数字个数代表多层感知机的隐藏层层数，每个数字的值代表多层感知机每层的神经元个数。
收敛阈值	参数范围为：[0.00000001, 0.000001]，默认值为：0.000001	收敛误差值，当损失函数取值优化到小于收敛阈值时停止迭代。
优化算法	<ul style="list-style-type: none"> l-bfgs: 有限内存中进行BFGS算法 gd: 梯度下降法 	解决函数最优化问题时使用的算法。
学习率	参数范围为：[0.00000001, 1]，默认为0.01	每次迭代的参数学习步长倍率
随机种子	任意整数，默认值为：1	随机初始化权重时的随机种子，随机种子设置为固定值时，每次结果是稳定的。
块大小	具体的取值和训练集的样本数目相关，不宜过大或过小，参数范围为：必须为大于等于32且小于等于10240的整数，默认为128。	训练时每次输入的样本块的大小，同一块的数据会在同一个矩阵中训练；
分类阈值	参数范围为：[0, 1]。默认值为：0.5	在分类中设置分类阈值thresholds。 如果模型预测结果为分类标签1的估计概率>thresholds，则预测为1，否则为0。高阈值是鼓励模型更频繁地预测0，反之则预测为1。 多分类请用英文逗号隔开，且数量与分类数相同，例如：分三类，示例：8, 9, 10

自动调参设置

系统将对设置指定或范围内的参数值循环调参，匹配出最优的组合。

🕒 自动调参设置
✕

拆分比例 * 评估标准

参数	指定值 *	范围 *	步数 *	是否使用指定值
优化算法	<input type="text" value="l-bfgs"/>	<input type="text" value="gd"/> + <input type="text" value="1"/>		<input type="checkbox"/>
最大迭代数	<input type="text" value="10"/>	<input type="text" value="10"/> - <input type="text" value="50"/>	<input type="text" value="2"/>	<input type="checkbox"/>
学习率	<input type="text" value="0.01"/>	<input type="text" value="0.01"/> - <input type="text" value="1"/>	<input type="text" value="2"/>	<input checked="" type="checkbox"/>
收敛阈值	<input type="text" value="0.000001"/>	<input type="text" value="1e-8"/> - <input type="text" value="0.000001"/>	<input type="text" value="2"/>	<input checked="" type="checkbox"/>
块大小	<input type="text" value="128"/>	<input type="text" value="128"/> - <input type="text" value="512"/>	<input type="text" value="2"/>	<input type="checkbox"/>

注意：勾选使用指定值时,不进行范围调参。

自动调参的方式分为两种：

- 指定值调参：指定一个固定的值进行自动调参。
- 范围调参：在指定的范围内进行自动调参。

设置项说明如下：

设置项	说明
拆分比例	将选择的数据拆分为两部分，一部分部分用于模型的评估，另一部分数据用于训练模型。
评估标准	用于选择数据的评估指标，包括：f1、precision、recall、accuracy、AUC(二分类)。 其中，评估标准“AUC(二分类)”仅对二分类问题生效。

参数		自动调参的参数项。 注意： <ul style="list-style-type: none"> • 优化算法的范围支持多选。 • 其他参数默认的范围提供了一个推荐值，并不是算法限制的最大值和最小值。
指定值调参	指定值	指定一个固定的值进行自动调参。需要先勾选“是否使用指定值”才能配置。
	是否使用指定值	控制是否使用使用指定值进行调参。
范围调参	范围	设置自动调参参数的范围。 若运行速度比较慢，可将参数范围调小一点。
	步数	进行范围调参时，在设置的范围内生成多少个参数值。 示例： 1) 范围为[3,5]，步数为3时，生成的参数值：3, 4, 5 2) 范围为[40,100]，步数为4时，生成的参数值：40, 60, 80, 100

示例

设置自动调参设置如图：

🕒 自动调参设置
✕

拆分比例 * 评估标准

参数	指定值 *	范围 *	步数 *	是否使用指定值
优化算法	<input type="text" value="l-bfgs"/>	<input type="text" value="l-bfgs"/>		<input type="checkbox"/>
最大迭代数	<input type="text" value="10"/>	<input type="text" value="10"/> - <input type="text" value="11"/>	<input type="text" value="6"/>	<input type="checkbox"/>
学习率	<input type="text" value="0.01"/>	<input type="text" value="0.01"/> - <input type="text" value="1"/>	<input type="text" value="2"/>	<input type="checkbox"/>
收敛阈值	<input type="text" value="0.000001"/>	<input type="text" value="0"/> - <input type="text" value="0.000001"/>	<input type="text" value="2"/>	<input type="checkbox"/>
块大小	<input type="text" value="128"/>	<input type="text" value="1"/> - <input type="text" value="1"/>	<input type="text" value="2"/>	<input type="checkbox"/>

注意：勾选使用指定值时,不进行范围调参。

在训练节点查看分析结果如图：

模型参数

超参数名称	超参数值
最大迭代数	10
感知机的隐藏层节点数量	6,5,6
收敛阈值	0.0
优化算法	l-bfgs
学习率	0.01
随机种子	1
块大小	1