

回归算法

回归分析是研究自变量和因变量之间关系的一种预测模型技术。这些技术应用于预测时间序列模型和找到变量之间关系。回归分析也是一种常用的统计学方法，经由统计机器学习融入机器学习领域。回归既可以指算法也可以指问题，因此在指代的时候容易混淆。实际上，回归就是一个过程而已。常用的回归算法包括

1. 普通最小二乘回归 (OLSR)
2. 线性回归
3. 逻辑回归
4. 逐步回归
5. 多元自适应回归样条法 (MARS)
6. 局部估计平滑散点图 (LOESS)

基于实例的学习方法

基于实例的学习通过训练数据的样本或实例建模，这些样本或实例也被视为建模所必需的。这类模型通过会建一个样本数据库，比较新的数据和数据库里的数据，通过这种方式找到最佳匹配并作出预测。换句话说，这类算法在做预测时，一般会使用相似度准则，比对待预测样本和原始样本之间的相似度，再做出预测。因此，基于实例的方法也被称之为赢家通吃的方法 (winner-take-all) 和基于记忆的学习 (memory-based learning)。常用的基于实例的学习算法包括：

1. k-邻近算法 kNN
2. 学习矢量量化算法 (LVQ)
3. 自组织映射算法 (SOM)
4. 局部加权学习算法 (LWL)

正则化算法

正则化算法背后的思路是，参数值较小的时候模型更加简单，对模型的复杂度会有一个惩罚值，偏好简单的、更容易泛化的模型，正则化算法可以说是这种方法的延伸。常用的正则化算法包括：

1. 岭回归
2. LASSO算法
3. Elastic Net
4. 最小角回归算法 (LARS)

决策树算法

决策树算法的目标是根据数据属性的实际值，创建一个预测样本目标值的模型。训练时，树状的结构会不断分叉，直到最初最终的巨册。也就是说，预测阶段模型会选择路径进行决策。决策树常被用于分类和回归。决策树一般速度快、结果准，因此也属于最受欢迎的机器学习算法之一。常用的决策树算法包括：

1. 分类和回归树 (CART)
2. ID3算法
3. C4.5算法和C5.0算法 (它们是一种算法的两种不同版本)
4. CHAID算法
5. 单层决策树

6. M5算法
7. 条件决策树

贝叶斯算法

贝叶斯方法指的是那些明确使用贝叶斯定理解决分类或回归等问题的算法。常用的贝叶斯算法包括：

1. 朴素贝叶斯算法
2. 高斯朴素贝叶斯算法
3. 多项式朴素贝叶斯算法
4. AODE算法
5. 贝叶斯信念网络（BBN）
6. 贝叶斯网络（BN）

聚类算法

聚类跟回归一样，既可以用来形容一类问题，可以指代一组方法。聚类方法通常涉及质心（centroid-based）或层次（hierarchal）等建模方式，所有的方式都与数据固有的结构有关，目标是将数据按照它们之间共性最大的组织方式分成几组。换句话说，算法将输入样本聚成围绕一些中心的数据团，通过这样的方式发现数据分布结构中的规律。常用的聚类算法包括：

1. K-均值
2. K-中位数
3. EM算法
4. 分层聚类算法

关联规则学习

关联规则学习再数据不同变量之间观察到了一些关联，算法要做的就是找出最能描述这些关系的规则，也就是获取一个事件和其他事件之间以来或关联的知识。常用的关联规则算法有：

1. Apriori算法
2. Eclat算法

人工神经网络

人工神经网络是一类受生物神经网络结构或功能启发而来的模型。它们是一类常用于解决回归和分类等问题的模式匹配，不过，实际上是一个含有成百上千种算法及各种问题变化的子集。注意这里将深度学习从人工神经网络算法中分离了出去，因为深度学习实在太受欢迎了。人工神经网络指的是更加经典的感知方法。常用的人工神经网络包括：

1. 感知机
2. 反向传播算法（BP神经网络）
3. Hopfield网络
4. 径向基函数网络（RBFN）

深度学习算法

深度学习算法是人工神经网络的升级版，充分利用廉价的算力。近年来，深度学习得到广泛应用，尤其是语音识别、图像识别。深度学习算法会搭建规模更大、结构更复杂的神经网络，正如上文所说，很多深度学习方法都涉及半监督学习问题，这种问题的数据一般量极大，而且只有很少部分带有标签。常用的深度学习算法包括：

1. 深度玻尔曼机 (DBM)
2. 深度信念网络 (DBN)
3. 卷积神经网络 (CNN)
4. 栈式自编码算法 (Stacked Auto-Encoder)

降维算法

降维算法和聚类有些类似，也是试图发现数据的固有结构。但是，降维算法采用的是无监督学习的方式，用更少（更低）的信息进行总结和描述。降维算法可以监督学习的方式，被用于多维数据的可视化或对数据进行简化处理。很多降维算法经过修改后，也被用于分类和回归的问题。常用的降维算法包括：

1. 主成分分析法 (PCA)
2. 主成分回归 (PCR)
3. 偏最小二乘回归 (PLSR)
4. 萨蒙映射
5. 多维尺度分析法 (MDS)
6. 投影寻踪法 (PP)
7. 线性判别分析法 (LDA)
8. 混合判别分析法 (MDA)
9. 二次判别分析法 (QDA)
10. 灵活判别分析法 (Flexible Discriminant Analysis, FDA)

模型融合算法

模型融合算法将多个简单的、分别单独训练的弱机器学习算法结合在一起，这些弱机器学习算法的预测以某种方式整合成一个预测。通常这个整合后的预测会比单独的预测更好一些。构建模型融合算法的主要精力一般用于决定将那些弱机器学习算法以什么样的方式结合在一起。模型融合算法是一类非常强大的算法，因此也很受欢迎。常用的模型融合增强方法包括：

1. Boosting
2. Bagging
3. AdaBoost
4. 堆叠泛化 (混合)
5. GBM算法
6. GBRT算法
7. 随机森林

其他

1. 特征选择算法
2. Algorithm accuracy evaluation
3. Performance measures

author zengbaocheng