算法综合分析：

**1、反向传导算法（BP）**

用处：求解神经网络的权重参数和偏置项，让神经网络的输出尽量和目标输出一样

原理：利用梯度下降的方法最小化均方差，将误差逐层传播并且修改权重和偏置项

优点：能够自我调节，使用于任意深的网络，是目前大量使用的方法

缺点：收敛速度慢，会出现局部最小化情况，会产生过拟合情况

**2、卷积神经网络（CNN）**

用处：CNN主要用来识别位移、缩放及其他形式扭曲不变性的二维图形，还可以用于音识别、人脸识别、通用物体识别、运动分析、自然语言处理

原理：CNN基本结构有两层，其一为特征提取层，每个神经元的输入与前一层的局部接受域相连，并提取该局部的特征。其二是特征映射层，网络的每个计算层由多个特征映射组成，每个特征映射是一个平面，平面上所有神经元的权值相等

优点：卷积神经网络中的每一个卷积层都紧跟着一个用来求局部平均与二次提取的计算层，这种特有的两次特征提取结构减小了特征分辨率。利用了共享权值，共享权值（卷积核）带来的直接好处是减少网络各层之间的连接，同时又降低了过拟合的风险。由于CNN的特征检测层通过训练数据进行学习，所以在使用CNN时，避免了显示的特征抽取，而隐式地从训练数据中进行学习；再者由于同一特征映射面上的神经元权值相同，所以网络可以并行学习，这也是卷积网络相对于神经元彼此相连网络的一大优势。卷积神经网络以其局部权值共享的特殊结构在语音识别和图像处理方面有着独特的优越性，其布局更接近于实际的生物神经网络，权值共享降低了网络的复杂性，特别是多维输入向量的图像可以直接输入网络这一特点避免了特征提取和分类过程中数据重建的复杂度。

缺点：构建模型非常复杂，训练时间比普通神经网络长

**3、PCA主成分分析**

用处：主成分分析（PCA）是一种能够极大提升无监督特征学习速度的数据降维算法

原理：数据变化的主方向就是协方差矩阵的主特征向量。选取协方差矩阵的主特征向量，舍去一些较小的向量，然后尝试还原近似的数据，得到的是降低了维度的数据

优点：PCA算法可以将输入向量转换为一个维数低很多的近似向量，而且误差非常

小。

缺点：前k个主特征向量的选取要多次试验，不然的话还原近似数据会相差太大

**4、Softmax回归**

用处：能够进行多元分类

原理：Logistic回归的推广，每一个类别都有一个函数来计算它的概率（而logistic只有两个类别0、1），权重矩阵是多个权重的排列，然后利用梯度下降法来求解权重矩阵

优点：扩展了logistic回归，可以对多个类别进行分类

缺点：类别要是互斥的，“冗余”性（冗余性指的是最优解不止一个，有多个）

**5、k最邻近算法**

用处：利用已有的数据来直接分类

原理：有一个样本数据集合，并且每个数据都含有标签，即知道每个数据与所属分类的对应关系。当输入新的没有标签的数据时，将新的数据的每个特征与样本集中数据相应特征进行比较，然后算法提取样本最相似的数据的分类标签。一般来说，只选取样本数据集中前K个最相似的数据，然后选取前K个数据中出现次数最多分类作为新数据的分类。

优点：无需经过复杂的训练就可以预测数据的类别；简单好用，容易理解，精度高，理论成熟，既可以用来做分类也可以用来做回归； 可用于数值型数据和离散型数据；训练时间复杂度为O(n)；无数据输入假定； 对异常值不敏感

*缺点：*   
计算复杂性高；空间复杂性高； 样本不平衡问题（即有些类别的样本数量很多，而其它样本的数量很少）； 一般数值很大的时候不用这个，计算量太大。但是单个样本又不能太少，否则容易发生误分；最大的缺点是无法给出数据的内在含义。

**6、SVM支持向量机**

用处：分类，从线性可分到线性不可分

原理：通过最大函数间隔来训练权重，然后用支持向量来区分数据，但是只适用于线性可分的数据。后来引入了核函数把线性不可分的数据映射到高维的空间实现线性可分

优点：可以解决高维问题，即大型特征空间；能够处理非线性特征的相互作用；无需依赖整个数据；可以提高泛化能力；

缺点：当观测样本很多时，效率并不是很高；对非线性问题没有通用解决方案，有时候很难找到一个合适的核函数；对缺失数据敏感；

**7、决策树**

用处：分类和回归的方法

原理：从树的根部开始，逐步按照节点的特征进行决策直至到达叶子，那么得到类别。构建：逐步按照信息熵增最大的特征进行构建决策树

优点：易于理解，好比人的决策过程一样。比较适合处理有缺失属性的样本；能够处理不相关的特征；在相对短的时间内能够对大型数据源做出可行且效果良好的结果

缺点：容易发生过拟合；忽略了数据之间的相关性；对于那些各类别样本数量不一致的数据，在决策树当中,信息增益的结果偏向于那些具有更多数值的特征

工具：

开源的深度学习框架比较好的是TensorFlow，其实就是相当于一个函数库，里面集成了一些常用的操作。TensorFlow 是一个采用数据流图（data flow graphs），用于数值计算的开源软件库。支持深度学习算法的构建。

英文官方网站：  
<http://tensorflow.org/>

官方GitHub仓库：  
<https://github.com/tensorflow/tensorflow>

中文版 GitHub 仓库：  
<https://github.com/jikexueyuanwiki/tensorflow-zh>