减少变量的关键点在于，如何确定新变量与原变量的关系，以及如何确定新变量的重要程度。主成分分析的实质是一种降维技术，它的实质就是确定原来变量在诸主成分上的荷载。对于此，我们采用主成分分析法，它的本质上是奇异值分解。我们将原始数据标准化。建立变量的相关系数阵求其的特征根及相应的单位特征向量。对于第一个问题，特征向量就是主成分新变量关于各个原变量的表达式的系数。于第二个问题，重要程度分别是相关矩阵较大的特征值所对应的特征向量。计算主成分贡献率及累计贡献率。取百分之80。

The key issue of reducing the variables is how to determine the relationship between the new variables and original variables, and how to define the importance of each new variables. The core of PCA is a technology to reduce the dimensionality, find the load of original variables to each new variable. We employ PCA, of which the essence is singular value decomposition. We standardize the original variables, set up the correlation coefficient matrix and solve the eigenvalues and eigenvectors. For the first problem, the eigenvectors are the coefficients of new PCs of expressions of each original variable. For the second problem, the importance is the contribution rates of eigenvalues. We calculate those of the first several large eigenvalues, taking 80%.

主成分回归就是在主成分分析的基础上进行回归。我们采用的是线性回归。虽然相关系数较高，但是无论何种模型必然会因为有些原自变量是离散的，有些是连续的，因此存在一定误差。更重要的，我们并不能直观的感受到到底何种模型是适合的。

PCR is to regress on the basis of PCA. The model we use is logistic model to show that the DV lies in the bound of [0,1). We calculate the correlation coefficients, which is higher than pure logistic

1701年托马斯贝叶斯提出的理论贝叶斯判别1701年贝叶斯公式

距离判别方法简单实用，但没有考虑到每个总体出现的机会大小，即先验概率，没有考虑到错判的损失。贝叶斯判别法正是为了解决这两个问题提出的判别分析方法。

通常，事件A在事件B(发生)的条件下的概率，与事件B在事件A的条件下的概率是不一样的；然而，这两者是有确定的关系,贝叶斯法则就是这种关系的陈述。

后验概率最大原则 平均误判最小原则 平均错判损失最小

现分别有高销量低销量两种手机，在好卖里分别有七个大清晰度手机和三个小清晰度手机，在不好卖里有一个大清晰度手机和九个小清晰度手机，现已知从所有手机里任意抽出了一个手机，问这个球是大清晰度且高销量概率是多少?假设已经抽出大清晰度为事件 B，选中高销量为事件 A，则有：P(B) = 8/20，P(A) = 1/2，P(B|A) = 7/10，按照公式，则有：P(A|B) = (7/10)\*(1/2) / (8/20) = 0.875

10/20 5/20 3/4

Alphago神经网络 图像识别 自动驾驶 科大讯飞

模拟人思考的方式，对每个输入设定权值。确定权重。小于阈值使用传递函数，大于阈值使用固定值。输入层 输出层 隐藏层。前一层输出作为后一层输入。类似于人的思考方式

非线性映射能力

能学习和存贮大量输入-输出模式映射关系，无需事先了解描述这种映射关系的数学方程。只要能提供足够多的样本模式供网络进行学习训练，它便能完成由n维输入空间到m维输出空间的非线性映射。

泛化能力

当向网络输入训练时未曾见过的非样本数据时，网络也能完成由输入空间向输出空间的正确映射。这种能力称为泛化能力。

容错能力

输入样本中带有较大的误差甚至个别错误对网络的输入输出规律影响很小。

We utilize Tangent Sigmoid function as the transfer function; we use Levenberg Marquardt algorithm (trainlm) as the training algorithm; we use the Gradient descent with momentum weight and bias learning function带动量项的BP学习规则(learngdm) as the learning algorithm; we use the mean square error (MSE) method as the learning function. 均方误差

XG Boosting成功预测特朗普上台 论文最新修改

多个专家的打分比一个专家好。汇集几个弱分类器成为强分类器。

公式的意义损失函数分类器与分类器有关的减少误差的函数

训练集：（A, 14岁），（B，16岁）、（C, 24岁）、（D, 26岁）；

训练数据的均值：20岁，（这个很重要，因为GBDT与i开始需要设置预测的均值，这样后面才会有残差！）

决策树的个数：2棵；

每个样本的特征有两个：购物金额是否小于1K；经常去百度提问还是回答；

下面开始GBDT学习

首先，输入初值20岁，根据第一个特征（具体选择哪些特征可以根据信息增益来计算选择），可以把4个样本分成两类，一类是购物金额<=1K，一类是>=1K的。假如这个时候我们就停止了第一棵树的学习，这时我们就可以统计一下每个叶子中包含哪些样本，这些样本的均值是多少，因为这个时候的均值就要作为所有被分到这个叶子的样本的预测值了。比如AB被分到左叶子，CD被分到右叶子，那么预测的结果就是：AB都是15岁，CD都是25岁。和他们的实际值一看，结果发现出现的残差，ABCD的残差分别是-1, 1, -1, 1。这个残差，我们要作为后面第二棵决策树的学习样本。

然后学习第二棵决策树，我们把第一棵的残差样本（A, -1岁）、（B，1岁）、（C, -1岁）、（D, 1岁）输入。此时我们选择的特征是经常去百度提问还是回答。这个时候我们又可以得到两部分，一部分是AC组成了左叶子，另一部分是BD组成的右叶子。那么，经过计算可知左叶子均值为-1，右叶子均值为1. 那么第二棵数的预测结果就是AC都是-1，BD都是1. 我们再来计算一下此时的残差，发现ABCD的残差都是0！停止学习~

这样，我们的两棵决策树就都学习好了。进入测试环节：

测试样本：请预测一个购物金额为3k，经常去百度问淘宝相关问题的女生的年龄~

我们提取2个特征：购物金额3k，经常去百度上面问问题；

第一棵树 —> 购物金额大于1k —> 右叶子，初步说明这个女生25岁

第二棵树 —> 经常去百度提问 —> 左叶子，说明这个女生的残差为-1；

叠加前面每棵树得到的结果：25-1=24岁，最终预测结果为24岁~