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%.

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 regression. However, we find a flaw that no matter which model we use, there exists a problem that some of the variables are discrete while some of those are continuous, which means we may increase the error by doing so. More importantly, we cannot know visually that which model is fit for the data.

The theory proposed by Thomas Bayes in 1701 and the Bayes formula.

Distance distinction is practical and easy to use, but it does not consider probability which each individual exists in, which is priori probability and it does not consider the loss caused by misjudgement. Bayes distinction is a method to solve the two issues in distinction.

In general, the probability that A happens when B happens, is different from the probability that B happens when A happens. However, the relationship between the two probabilities are definite, which comes to Bayes rule.

Posterior Probability Maximum Principle

The Principle of Minimum Average Misjudgement

Reduce the average misjudgement loss

We have high sales and low sales, two kinds of phones. In high sales we have 7 high display resolution and 3 low display resolution. In low sales we have 1 high display resolution and 9 low display resolution. We randomly pick out a phone in all the phones, find the probability that the phone is high display resolution and high sales

Assume picking out high display resolution is event B, picking out high sales is event A, then we have P(B) = 8/20，P(A) = 1/2，P(B|A) = 7/10，according to the formula, we have：P(A|B) = (7/10)\*(1/2) / (8/20) = 0.875.

Alpha neural network image recognition autopilot control machine translation

Imitate the way that a person thinks in, set weights to each input. If it is lower than threshold, then we use transfer function. Larger we use fixed value. Input/output/hidden layer. The input of former layer is used as the input of latter layer. Mimic the way that a person thinks.

Nonlinear mapping capability

It can save large mapping function and we do not need to preset a model of the supervision learning method. As long as enough samples are given, it can generate a non-linear mapping from the IVs to the DVs.

Generalization ability

When the model trains the sample that has not seen yet, the Neural network can finish a mapping from input to output without too much error

fault-tolerant ability

If the input has a few error, even erroneous data, it has little impact on the output.

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

XG Boosting successfully predicted president Trump coming into power

The scores of multiple experts are more accurate than those of a sole expert. Synthesize several weak classifier into a strong classifier.

The meaning of the formula loss function classifier a function related to classifier to reduce error

Training set

The algebra mean of the training data: 20 years old

Two dimensions for dicision

Two traits of each sample

Residual as the sample of the second decision tree.