It can date back to 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: rather than the probability on the ensemble

The Principle of Minimum Average Misjudgement: reduce the occurrence of the misjudgement

Reduce the average misjudgement loss: reduce the loss of misjudgment when it happens

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).

The output is discrete rather than continuous.

Alpha 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

The rise of XG Boosting is that it 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

Here comes the example. We have four people, and we want to predict the age of the people through traits. We establish two decision trees, which are two traits. First we divide people according to first trait. The algebra means of the same category of the training data are 15 years old and 25 years old. We then compute the residual, -1,1,-1,1 as the training set of the second traits, and repeat the process. The residual is 0, which means we finish the training part. If we have a person whose Monthly shopping Amount is 3000 yuan and often ask question of Quora, we can predict the age is 25 -1=24.

As the score shown in the table, it is relatively higher.