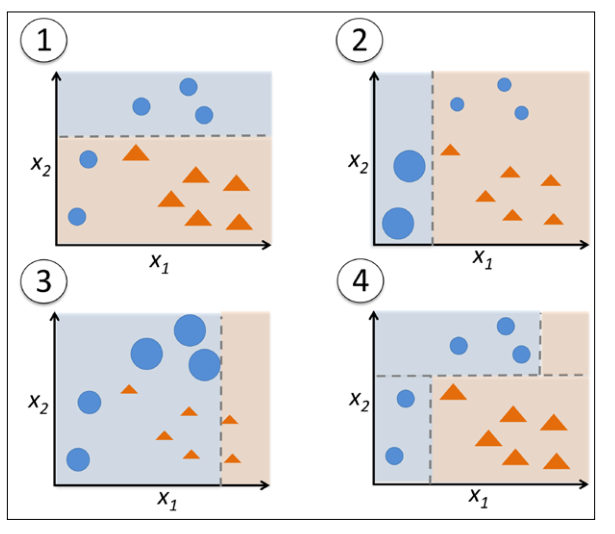
提升算法从弱学习算法出发，反复学习，得到一系列弱分类器（又称为基本分类器），然后组合这些弱分类器，构成一个强分类器。

A typical example of a weak learner would be a decision tree stump. The key concept behind boosting is to focus on training samples that are hard to classify, that is , to let the weak learners subsequently learn from misclassified training samples to improve the performance of the ensemble.

The original boosting procedure is summarized in four key steps as follows:

1. Draw a random subset of training samples d1 without replacement from the training set D to training a weak learner C1
2. Draw second random training subset d2 without replacement from the training set and add 50 percent of the samples that were previously misclassified to train a weak learner C2
3. Find the training samples d3 in the training set D on which C1 and C2 disagree to train a third weak learner C3
4. Combine the weak learner C1,C2, and C3 via majority voting.

AdaBoost uses the complete training set to train the weak learners where the training samples are reweighted in each iteration to build a strong classifier that learns from the mistakes of the previous weak learners in the ensemble.

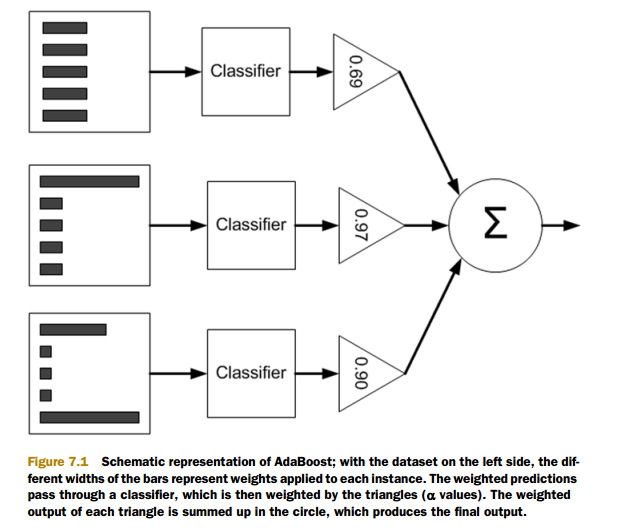


Subfigure1 : represents a training set for binary classification where all training samples are assigned equal weights. Based on this training set, we train a decision stump (shown as a dashed line) that tries to classify the samples of the two classes (triangles and circles) as well as possible by minimizing the cost function (or the impurity score in the special case of decision tree ensembles).

Subfigure2: we assign a larger weight to the two previously misclassifed samples (circles). Furthermore, we lower the weight of the correctly classifed samples. The next decision stump will now be more focused on the training samples that have the largest weights, that is, the training samples that are supposedly hard to classify. The weak learner shown in subfgure 2 misclassifes three different samples from the circle-class, which are then assigned a larger weight as shown in subfgure 3.

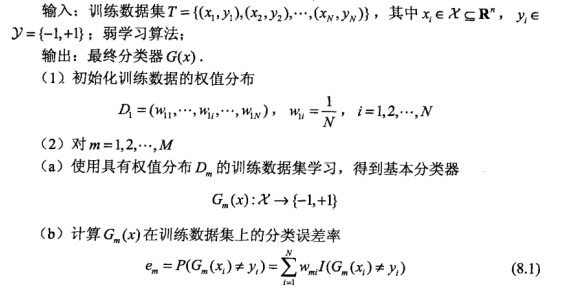
Assuming that our AdaBoost ensemble only consists of three rounds of boosting, we would then combine the three weak learners trained on different reweighted training subsets by a weighted majority vote, as shown in subfgure 4.

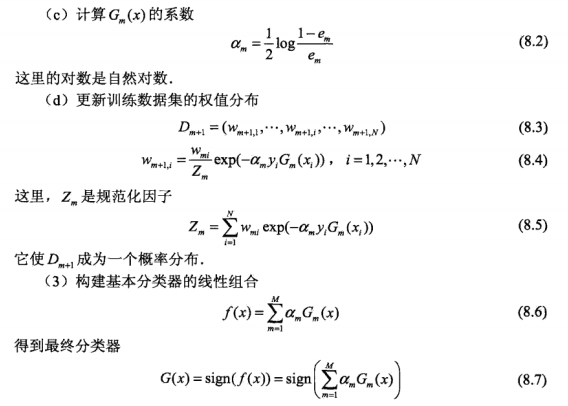
下图对AdaBoost 进行了解释：



每个数据点的权重一开始是相等的。在后续的迭代中，AdaBoost 算法会在增加错分数据的权重的同时，降低正确分类数据的权重。D是一个概率分布向量，因此其所有的元素之和为1.0.

AdaBoost算法步骤：（注：统计机器学习和Machine Learning in Action 中的公式有些不同，但整体意思差不多）





Reference:

1. python machine learning
2. 统计机器学习
3. Machine learning in action