

Techniques for Analyzing Stochastic Time-Series Data

Dennis Castleberry Brandon Oubre Haikou Yu

November 19, 2013

The Naive Bayes Classifier

- Reduce classification to probability. What is $P(class|attribute1, attribute2, \dots, attributeN)$.
- Assumes that each attribute is independent of the others. (Hence the “Naive” nickname.)
- For example, let's consider if a car is stolen using $P(stolen|Color, Type)$. Naive Bayes will assume $color = red$ and $type = sportscar$ to be independent.
- Naive Bayes is not sensitive to irrelevant attributes, since the probabilities of such attributes will be similar for all classes.
- Naive Bayes is quick to train, as it requires only one pass-through of the training data.

Naive Bayes in Action

Training Data			
Over 170cm	Eye Color	Hair Length	Sex
No	Blue	Short	Male
Yes	Brown	Long	Female
No	Blue	Long	Female
Yes	Brown	Short	Male
Yes	Brown	Short	Female

Only discrete values shown, but we can still interpret real data using normal distributions!

Suppose we are given an unseen data point $\langle \text{No}, \text{Blue}, \text{Short} \rangle$.
What should we classify it as?

Naive Bayes in Action

$$\begin{aligned} &P(\text{Male}|\text{No}, \text{Blue}, \text{Short}) \\ &= \frac{P(\text{No}, \text{Blue}, \text{Short}|\text{Male})P(\text{Male})}{P(\text{No}, \text{Blue}, \text{Short})} \\ &= \alpha P(\text{Male})P(\text{No}|\text{Male})P(\text{Blue}|\text{Male})P(\text{Short}|\text{Male}) \\ &= \alpha \times \frac{2}{5} \times \frac{1}{2} \times \frac{1}{2} \times \frac{2}{2} = \boxed{0.1\alpha} \end{aligned}$$

$$\begin{aligned} &P(\text{Female}|\text{No}, \text{Blue}, \text{Short}) \\ &= \alpha P(\text{Female})P(\text{No}|\text{Female})P(\text{Blue}|\text{Female})P(\text{Short}|\text{Female}) \\ &= \alpha \times \frac{3}{5} \times \frac{1}{3} \times \frac{1}{3} \times \frac{1}{3} = \boxed{0.0\bar{2}\alpha} \end{aligned}$$

Since $P(\text{Male}|\text{Data}) > P(\text{Female}|\text{Data})$, we classify the unseen point as Male. For multiple classes, just select the class with the greatest probability!

Support Vector Machines (SVM)

- Idea is to draw a line (or hyperplane) between the data points of different classes. Classify unseen data by testing which side of the line it is on.
- Focus on support vectors, or the points that would change the line if removed from the training data.
- Find an optimal line to separate the data. Such a line will have the larger margin for data points and should mis-classify the least number of new points.
- If data is not linearly separable, then a transformation of the data to a new basis can be performed. The data may be linearly separable in the new basis.

SVM Example

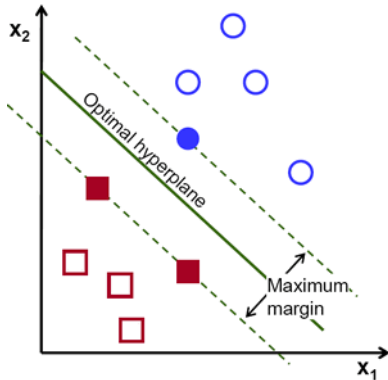


Image from http://docs.opencv.org/doc/tutorials/ml/introduction_to_svm/introduction_to_svm.html

- Solid Figures are support vectors.
- Due to the maximized margin, unseen figures can be closer to the line than the support vectors and still be correctly classified.
- It is easy to see how new points are classified.

Neural Networks

- Inspired by biological neurons.
- Neurons maintain a weighted sum of their inputs. The result of this sum is passed into a function and output. (A step function produces on/off signals while a Sigmoid will produce continuous levels of activation.)
- The network can be trained by adjusting the weights of the inputs to each neuron.
- In a feed-forward network, the backwards propagation algorithm accomplishes this.
- Networks with multiple layers can classify various types of non-linearly separable data.

An Artificial Neuron

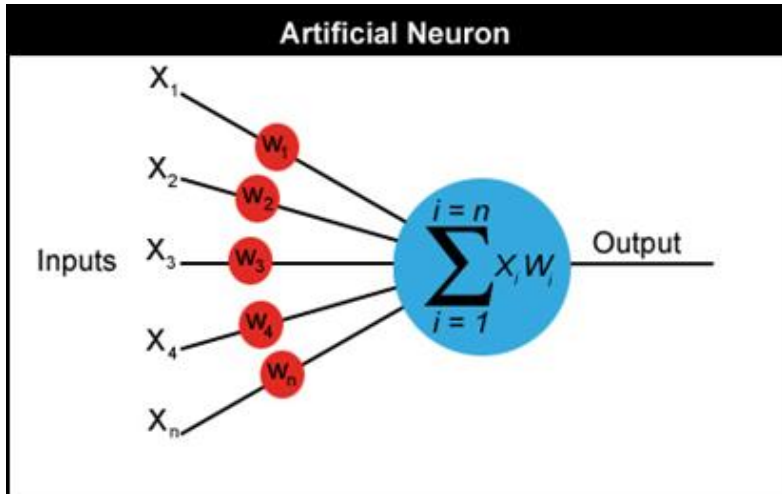


Image from <http://www.ai-junkie.com/ann/evolved/nnt1.html>

Neural Network Classification

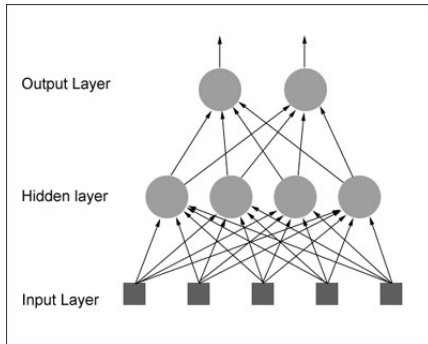


Image from <http://www.ai-junkie.com/ann/evolved/nnt1.html>

- Information is fed into the input layer.
- The outputs of the neurons in the output layer represent classifications.
- Hidden layers perform intermediary manipulations of signals. More hidden layers can be added as needed.