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 $P(class|attribute1, attribute2, ..., attributeN)$.
- 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-though 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 $\langle No, Blue, Short \rangle$. What should we classify it as?

Naive Bayes in Action

$$P(Male|No, Blue, Short)$$
$$= \frac{P(No, Blue, Short|Male)P(Male)}{P(No, Blue, Short)}$$
$$= \alpha P(Male)\boldsymbol{P(No|Male)P(Blue|Male)P(Short|Male)}$$
$$= \alpha \times \frac{2}{5} \times \frac{1}{2} \times \frac{1}{2} \times \frac{2}{2} = \boxed{0.1\alpha}$$

$$P(Female|No, Blue, Short)$$
$$= \alpha P(Female)P(No|Female)P(Blue|Female)P(Short|Female)$$
$$= \alpha \times \frac{3}{5} \times \frac{1}{3} \times \frac{1}{3} \times \frac{1}{3} = \boxed{0.0\overline{2}\alpha}$$

Since $P(Male|Data) > P(Female|Data)$, 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

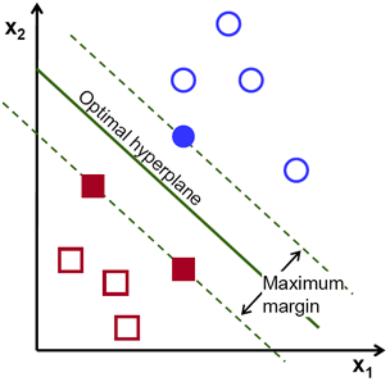

Image from http://docs.opencv.org/doc/tutorials/ml/introduction_to_svm/introduction_to_svm.html

- 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