

Lecture 8: Classifiers

Intro to Data Science for Public Policy, Spring 2016

by Jeff Chen & Dan Hammer, Georgetown University McCourt School of Public Policy

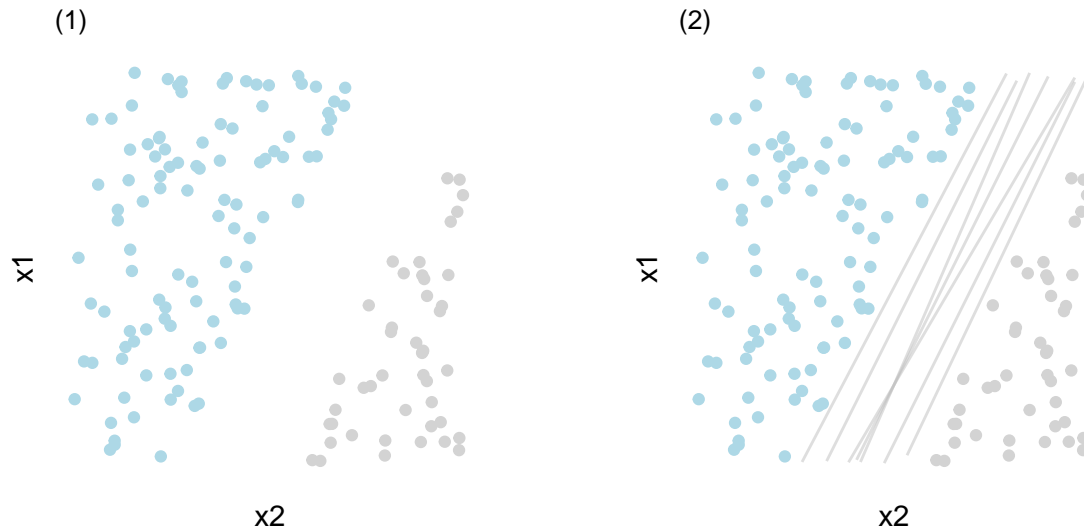
Contents

.....	1
Support vector machines	1
Artificial Neural Networks	2
Applications of classifiers	2

Building on Lecture 7, this section lightly introduces three more classification algorithms: Generalized Linear Models, Support Vector Machines, and Artificial Neural Networks.

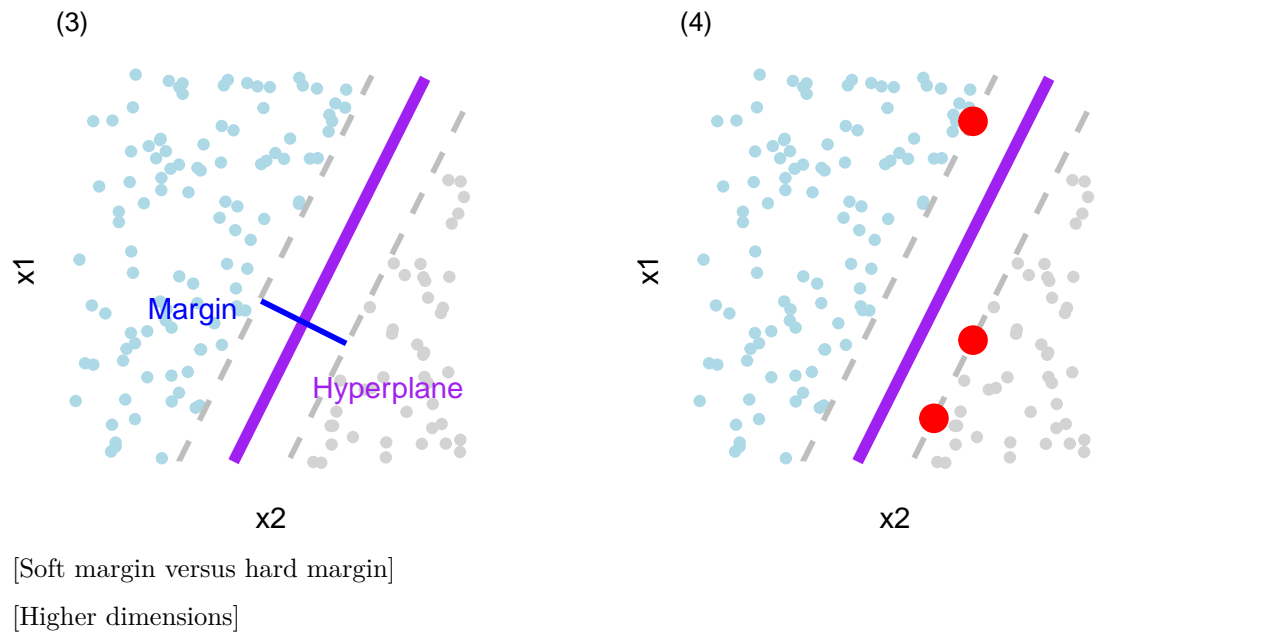
Support vector machines

You're given a three feature data set. Two features are continuous inputs and the third is a target containing labels of two groups. Upon plotting the points in two dimensional space and color coding for the target, you notice that there is a clear line of separation: A straight line can partition one group from the other (Figure 1). You think and realize that multiple lines could do the job: there are almost infinite lines (Figure 2) that could serve as the boundary between the groups. But which is the best? There should in theory be one line that optimally describes the separation between the groups.



If we are to assume a straight line is appropriate, we can find a line that maximizes the distance between the groups. As seen in Figure 3, the line is known as the hyperplane or decision boundary (in purple) and the distance is known as the margin (in blue). This is the fundamental goal of Support Vector Machines (SVM): to maximize the margin around a hyperplane. By maximizing the margin around the hyperplane, we also improve the ability for the model to tolerate mistakes that cross into the margin.

The algorithm looks for points that lie on the edges of the margin to act as “support vectors” of the hyperplane. Otherwise stated, these are control points or reference points that serve as markers that indicate the edges of the margin.



Artificial Neural Networks

Applications of classifiers

Appropriate uses of classification techniques

[text goes here]

#

Scoring

[text goes here]

#

prediction and prioritization

[text goes here]

#

Propensity score matching

[text goes here]

#

Exercise Data

- [Labor and wage analysis]