Support Vector Classification and Regression Models

Predictive Analytics

Jesús Calderón

Learning Objectives

By the end of this week, students will be able to:

- Describe and explain Support Vector Machine models.
- Discuss the advantages and limitations of these approaches.

Support Vector Machines

Nonlinear Transforamtions of the Feature Space

- Linear models are limited in the decision boundaries that they can produce.
- Support Vector Machines (SVM) use linear models to implement nonlinear boundaries by transforming the feature space into a new space using nonlinear functions.
- A linear model in the transformed space can represent nonlinear decision boundaries in the original space.

- For example, say we have two features, a_1 and a_2. We can create a new feature x:

$$x = w_1 a_1^3 + w_2 a_1^2 a_2 + w_3 a_1 a_2^2 + w_4 a_2^3$$

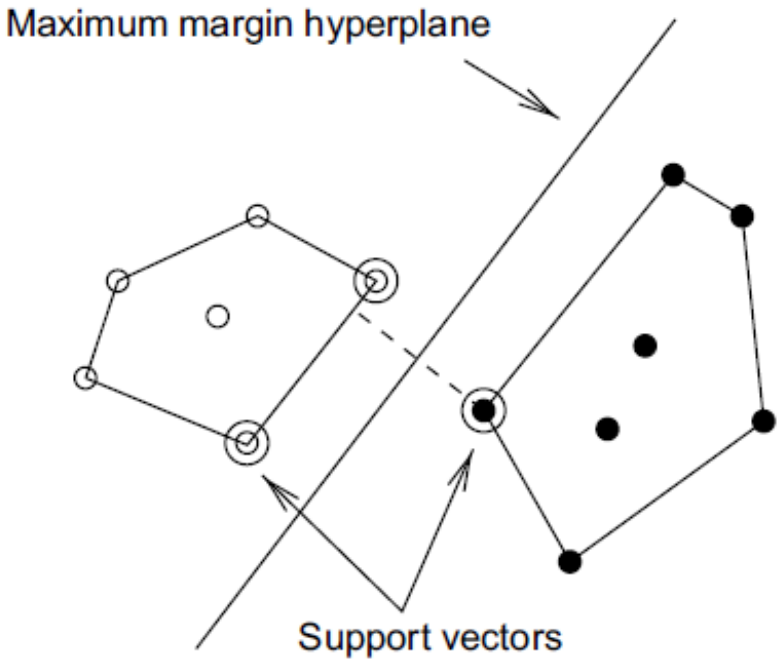
- This transformation may be promising given that polynomials of sufficiently high degree can approximate arbitrary decision boundaries to any required accuracy.
- The learning problem would now also require for us to estimate w_1, …, w_2.

Support Vector Machines

- If we simply transformed the input space as in the example and tried to apply a linear model, we would face to issues:
  - Computational complexity: with 10 attributes in the original data set, if we wanted to include all products with five factors, the learning algorithm would need to learn 2000 coefficients.
  - Overfitting: if the number of coefficients is large, the resulting model will overfit the training data.
- SVM try to address both problems, computational complexity and overfitting, by finding the *maximum margin hyperplane*.

- The maximum margin hyperplane is the one that gives the greatest separation between the classes.
- The *convex hull* of a set of points is the tightest enclosing convex polygon: it is the "outline" of all the points.
- Among all hyperplanes that could separate classes, the maximum margin hyperplane is the one that is as far as possible from both convex hulls.
- It is also the perpendicular bisector of the shortest line connecting the hulls.

Maximum Margin Hyperplane and Support Vectors



Maximum Margin Hyperplane (Witten et al, 2017)

- The instances closest to the maximum margin hyperplane (the ones closest to it) are called *support vectors*.
- There is always at least one support vector, but many times there are more.
- One can always reconstruct the maximum margin hyperplane based soleely on the support vectors, therefore, we do not need any of the other observations once the support vectors are determined.

Pros and Cons of SVM

- Overfitting is reduced.
- The maximum margin hyperplane is relatively stable: it does not change when the inputs change.
- Computational complexity is addressed by observing that many operations that would be required in the transformed space can actually be computed in the original space.
- Other nonlinear transformations are:
  - Radia Basis Function (RBF)
  - Sigmoid Function

References

References

- James, G., D. Witten, T. Hastie, and R. Tibshirani. *An Introduction to Statistical Learning with Applications in R*. United States: Springer, 2017.
- Witten, F., E. Frank, M. Hall, C. Pal. *Data Mining: Practical Machine Learning Tools and Techniques*. United States: Morgan Kaufmann, 2017. 4th Edition.