Melissa LaHoud

Grand Canyon University

DSC 540: Machine Learning

Dr. Aiman Darwiche

4/14/2021

Number of Misclassifications vs Maximizing the Margin

You have a choice of handling a binary classification task using *number of misclassifications* as the performance measure and maximizing the margin between the two classes as the performance measure. On what factors does your decision depend? Provide a formal explanation, supported by theorems and ideas presented in the readings associated with this topic.

If we look at support vector machine for a binary classification, we know that support vector machine algorithm is to find a hyperplane in N-dimensional space that distinctly classifies the data points. To separate the two classes of data points, there are many possible hyperplanes that could be chosen. Our objective is to find a plane that has the maximum margin, i.e the maximum distance between data points of both classes. Maximizing the margin distance provides some reinforcement so that future data points can be classified with more confidence.

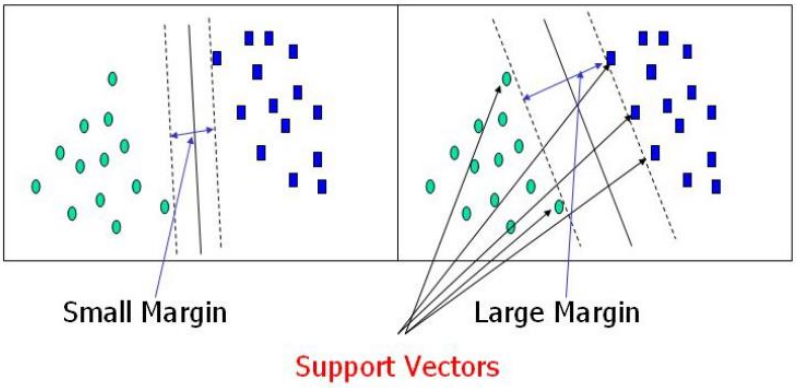

Figure 1

Support vectors are data points that are closer to the hyperplane and influence the position and orientation of the hyperplane. Using these support vectors, we maximize the margin of the classifier. Deleting the support vectors will change the position of the hyperplane. These are the points that help us build our SVM. (Gandhi, R., 2018) SVM cannot find a separating

hyperplane that minimizes the number of misclassified data points. SVM, in this case, is not searching for the hard margin, which will classify all data flawlessly. Instead, SVM is now a soft-margin classifier; that is, SVM is classifying most of the data correctly, while allowing the model to misclassify a few points in the vicinity of the separating boundary. We look at minimization of the objective function.

$$J(w,b,\xi)=\frac{1}{2}\|w\|^2+C\sum_{i=1}^{N}\xi_i,$$

"The regularization term or box constraint, C, is a parameter that varies, depending on the optimization goal. As C is increased, a tighter margin is obtained, and more emphasis is placed on minimizing the number of misclassifications. As C is decreased, more violations are allowed, because maximizing the margin between the two classes becomes the SVM aim. Figure 2 captures the effect of the regularization parameter, with respect to margin width and misclassification. For C1<C2, fewer training points are within the margin for C2 than for C1, but the latter has a wider margin.

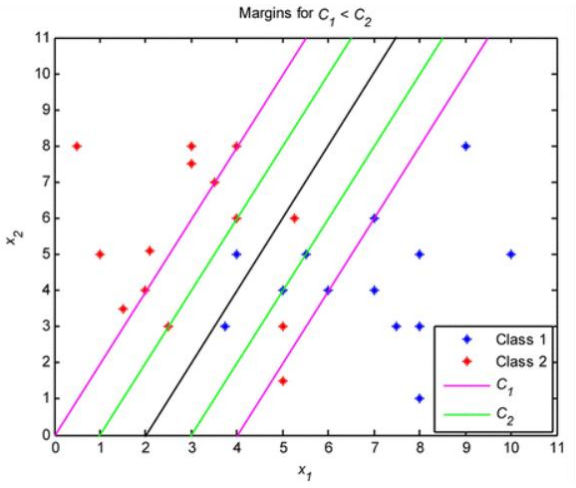

Figure 2

Therefore, I would choose maximizing the margin between the two classes as the performance measure over using number of misclassifications as the performance measure due to the fact that a method such as SVM allows misclassifications, and the misclassifications depend on how large the margins are as well. Also, Minimizing the number classifications causes overfitting as well which happens when a model learns the detail and noise in the training data to the extent that it negatively impacts the performance of the model on new data. This means that the noise or random fluctuations in the training data is picked up and learned as concepts by the model. The problem is that these concepts do not apply to new data and negatively impact the model's ability to generalize. (Brownlee, J., 2019)

Linear SVM vs Perceptron Algorithm

You have a choice of handling a binary classification task using (i) *linear SVM*, and (ii) *perceptron algorithm*. On what factors does your decision depend? Provide a formal explanation, supported by theorems and ideas presented in the readings associated with this topic.

First, I think it's important to know what each of the choices are. Perceptron learning is one of the most primitive form of learning and it is used to classify linearly separable datasets. Additionally, combination of perceptron's can also be used to model the non-linearity of the data. Given two data points, Perceptron tries to find a hyperplane that separates the data. There can be numerous hyperplane that separates two data points in Perceptron. The Support Vector Machine is a classification method. The SVM classifier is widely used in bioinformatics due to its high accuracy, ability to deal with high-dimensional data such as gene expression, and flexibility in modeling diverse sources of data. A support vector machine is a supervised

machine learning model that uses classification algorithms for two-group classification problems. After giving an SVM model sets of labeled training data for each category, they're able to categorize new datasets with some degree of accuracy depending upon the training datasets. Given two data points, support vector tries to find a hyperplane that maximizes the margin between those data points. (Kharel, S., 2020)

The condition of stoppage of algorithm are the key difference between these two. Perceptron stops after it classifies data correctly whereas SVM stops after finding the best plane that has the maximum margin, i.e. the maximum distance between data points of both classes. Maximizing the margin distance provides some reinforcement so that future data points can be classified with more confidence. Therefore, support vector machine is best suited for a binary classification task as it is less complex and can classify the data points based on calculating the width of the decision boundary between the two classes. SVM keeps a classification margin on each side so that it classifies test data points that come near to the boundary properly. The margin of the SVM makes SVM more robust in getting more closer to the real boundary (target function) of the datasets.

References

Brownlee, J. (2019, August 12). Overfitting and Underfitting With Machine Learning Algorithms. Retrieved from https://machinelearningmastery.com/overfitting-and-underfitting-with-machine-learning-algorithms/

Gandhi, R. (2018, June 7). Support Vector Machine — Introduction to Machine Learning Algorithms. Retrieved from https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47

Kharel, S. (2020, May 13). Perceptron vs SVM: A quick comparison. Retrieved from https://medium.com/@subashkharel/perceptron-vs-svm-a-quick-comparison-6b5d6b5d64f#:~:text=Perceptron learning algorithm works better,can generate in different experiments.&text=SVM keeps a classification margin,near to the boundary properly.