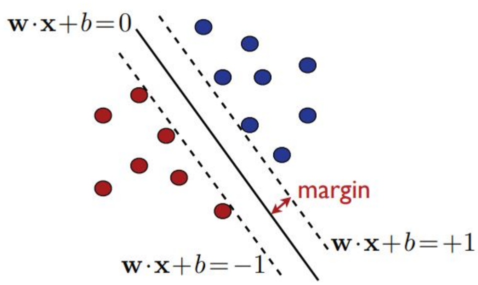
SUPPORT VECTOR MACHINES

1. Introduction. Machine Learning is the study that uses data to predict outcomes of unknown processes. The algorithms learn the relationships between input variables (features) by studying the input (training) data, and produce the output in the form of a weight vector. The weights are assigned to each feature based on its influence on the output. These outcomes can either be continuous, such as length, weight, etc, or categorical, such as "disease" and "no disease." A machine learning model that predicts the outcome of a process whose outcome is of categorical nature, is called a **Classification** model. Similarly, a **Regression** model predicts continuous values for other processes.

Support Vector Machines (SVMs) are a particularly powerful and general purpose class of supervised learning algorithms for both classification and regression problems. Given data, the algorithm outputs an optimal hyperplane (a linear predictor) which classifies new examples into categories.The SVM solution is the separating hyperplane with the maximum geometric margin and is thus known as the maximum-margin hyperplane.[2] SVM has usage in image processing, pattern, speech and emotion recognition, etc. I'm interested in a machine learning implementation of SVM for facial recognition.



What is a hyperplane? In a d-dimensional space, (usually R^d) a hyperplane is a linear affine subspace of $d-1$ dimensions. In a two dimensional space, this hyperplane is a line dividing the plane into two parts, each part belonging to a different class. For example, in a two dimensional space (R^2), if we intend to classify all non-zero real numbers as positive or negative, the line $x = 0$ is the required hyperplane.

Limitation: Training time, i.e. time taken to develop the output hypothesis is long for large datasets.

---

2. SVD. For the purpose of image recognition, the input data will be images of say [62×47] size. Then, each image has nearly 3,000 pixels. We could simply proceed by using each feature, but it is often more effective to pre-process the dataset to extract meaningful features. We will use a principal component analysis to extract the fundamental components to feed into our support vector machine classifier.

Principal Component Analysis (PCA) is essentially singular value decomposition, and will help us generate low-rank approximations to matrices. Think of these singular vectors (principal components) as linear combinations of the features and observed values that capture the essence, the most important characteristics in a least-squares sense.

---

3. Point-Line Duality. Duality relates to the inversion of a maximization problem into a minimization problem, or vice-versa, through a change of variables. Consider a point (a,b) on the Cartesian Plane. It's dual is the line y = ax - b. The solution to the original (primal) problem of minimizing a function f(x) w.r.t. x is bounded below by the solution of the dual problem that maximises an other function g w.r.t an other variable λ.

f(x) will be the error function, which tells us how far off our prediction is from the actual output value. Its optimum will be determined by taking the gradient of the dual problem and equating it to 0, which leads us to the advanced calculus concept.

---

4. Optimization. Gradient descent is an iterative optimization procedure in which at each step we improve the solution by taking a step along the negative of the gradient of the function to be minimized at the current point. [3] Imagine a bowl-shaped loss function. It is our intention to move towards the bottom of the bowl, which is where the height of the bowl (the value of the loss function) is minimum.

The simplest approach to using gradient information is to choose the weight update to comprise a small step in the direction of the negative gradient. After each such update, the gradient is re-evaluated for the new weight vector and the process repeated.[1]

REFERENCES

[1] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer Science+Business Media, LLC, 233 Spring Street, New York, NY 10013, USA, 2006.
[2] T. Mohri, Rostamizadeh, *Foundations of Machine Learning*, The MIT Press, Cambridge, Massachusetts, 2 ed., 2018.
[3] S. B.-D. Shai Shalev-Shwartz, *Understanding Machine Learning: From Theory to Algorithms*, Cambridge University Press, Cambridge, UK, 2014.