Support Vector Machine (SVM)

A log-barrier approach

Anurag[1]

[1]Research Scholar
Electrical Engineering
Indian Institute of Science, Bangalore

Convex optimization and applications
2 May, 2018

anuragg.in@gmail.com

Introduction

Convex optimization

$$\min_{x \in D} f(x) \tag{1}$$

where $f(x)$ and D are convex.

Quadratic program

$$\min_{Ax \leq b} x^T P x \tag{2}$$

where $P \succeq 0$ for a convex program.

Log barrier method

Aim

To make inequality constraint implicit in the objective function.

Indicator function

$$I_{-}(u) = \begin{cases} 0; u \leq 0 \\ \infty; u > 0 \end{cases} \tag{3}$$

Approximate indicator function (Log barrier function)

$$\hat{I}_{-}(u) = -(1/t)log(-u) \tag{4}$$

$$dom(\hat{I}_{-}) = -\mathbb{R}_{++}$$

anuragg.in@gmail.com

Approximate log barrier function

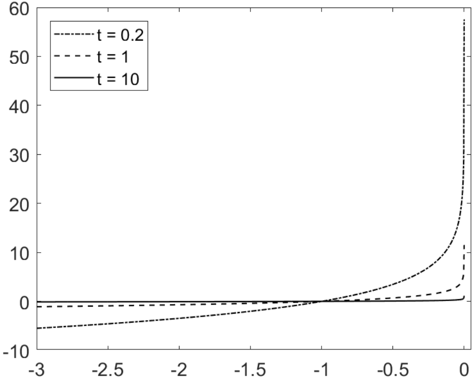

Figure 1: Log barrier function for t = 0.2, 1, 10

Support Vector Machine

Given a set of linearly separable data points, find a separting hyperplane that maximizes the margin between itself and the nearest data points.

(Definition) Margin

Twice of minimum distance between the separating hyperplane and data points.

Support Vector Machine

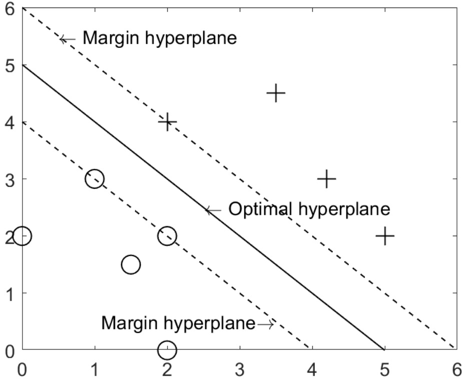

Figure 2: Separating hyperplane for linearly separable data points

Cost function for SVM

Given a set of linearly separable data points and labels (x_i, y_i), finding the maximum margin separating hyperplane is equivalent to the constrained problem:

Quadratic program for SVM

$$\begin{aligned} &\min_{w,b} \|w\|^2 \\ &s.t.\ y_i(\langle w, x_i \rangle + b) \geq 1,\ \forall i \end{aligned} \tag{5}$$

where $\langle w, x \rangle + b = 0$ is the equation of hyperplane (decision boundary).

SVM using log-barrier method

Quadratic program for SVM

$$\min_{w,b} \|w\|^2$$
$$s.t.\ y_i(\langle w, x_i \rangle + b) \geq 1,\ \forall i \tag{6}$$

Optimization problem using log barrier method

$$\min_{w,b} \left[\|w\|^2 - (1/t) \sum_i log(-1 + y_i(\langle w, x_i \rangle + b)) \right] \tag{7}$$

$$where\ t \in \mathbb{R}_+$$

Feasible start point

- Presence of logarithm in objective function restricts start of optimization from any random initial point
- Need for additional mechanism to find a feasible starting point

Optimization problem for a feasible start point

$$\min_{w,b,s} s$$
$$s.t.\ (1 - y_i(\langle w, x_i \rangle + b)) \leq s;\ \forall i \tag{8}$$

- Can put a lower bound on s to prevent unnecessary computation

Algorithm

- Fetch the dataset
- Identify a feasible starting point for SVM model (using a log barrier method) (7) using (8)
- Run the optimization problem (7) with the obtained starting point

Result

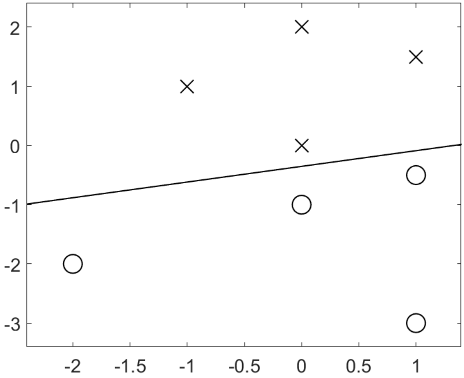

Figure 3: Separating hyperplane for a linearly separable 2D dataset using log barrier method (t = 1)

Result

Difficult to minimize the optimization problem (7) for large value of t in one step since its Hessian varies rapidly near the boundary of the feasible set.

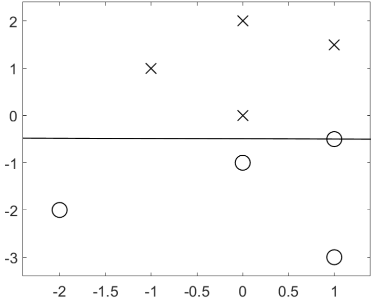

Figure 4: Separating hyperplane for a linearly separable 2D dataset using log barrier method ($t = 10$)

anuragg.in@gmail.com

Central path

(Defintion) Central path

Let $x^*(t),\ t > 0$ be the solution of

$$\begin{aligned} \min\ & t(f_0(x)) + \phi(x) \\ s.t.\ & Ax = b \end{aligned} \tag{9}$$

Then the central path associated with it is defined as the set of points $x^*(t),\ t > 0$ which we call the set central points.

anuragg.in@gmail.com

Result

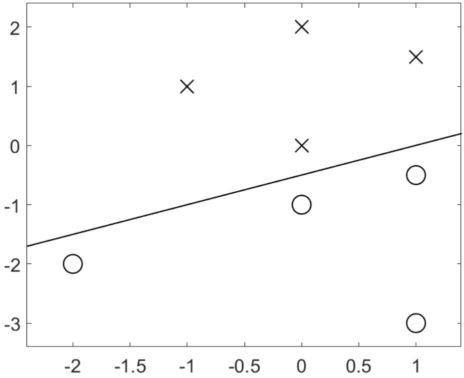

Figure 5: Separating hyperplane for a linearly separable 2D dataset using log barrier method ($t = 10$) by incorporating the concept of central path

Result

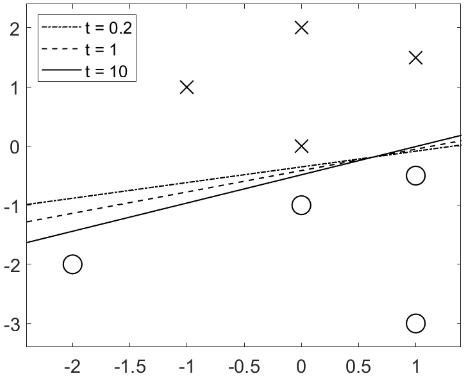

Figure 6: Central path of separating hyperplane (t = 0.2, 1, 10) for a linearly separable dataset

Comparison

Exact method

$$\begin{bmatrix}-1 & 2\end{bmatrix}\begin{bmatrix}x\\ y\end{bmatrix}+1=0$$

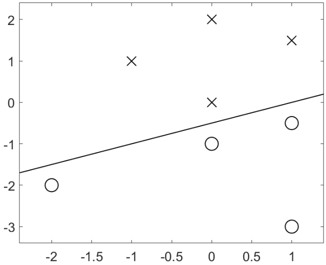

Figure 7: SVM using exact method

Log barrier method

$$\begin{bmatrix}-1.0024 & 2.0994\end{bmatrix}\begin{bmatrix}x\\ y\end{bmatrix}+1.0204=0$$

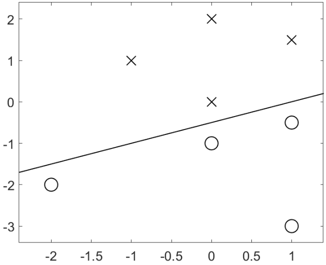

Figure 8: SVM using log barrier method and central path (t = 10)

anuragg.in@gmail.com

Soft margin SVM (Motivation)

When a dataset is not linearly separable due to insufficient features, noise and spurious data.

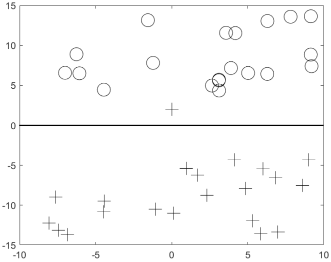

Figure 9: Soft margin to increase margin

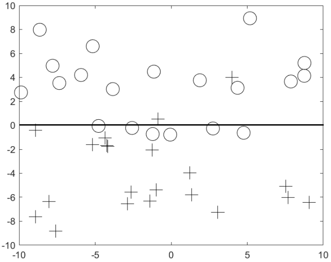

Figure 10: Soft margin to handle linearly inseparable dataset

Soft margin SVM

Soft margin SVM classification

$$\min_{\substack{w,b,\\ \xi_i \in \mathbb{R}_+}} \left[\|w\|^2 + C \sum_i^n \xi_i \right] \tag{10}$$
$$s.t.\ [1 - y_i(\langle w, x_i \rangle + b)] \leq \xi_i;\ \forall i$$

where C is a regularization/penalty parameter

Now, we have another set of inequalities that are introduced by non-negativity condition on ξ_i , $\forall i$. This has to be separately handled by another log barrier function.

Soft margin SVM using log barrier method

Soft margin SVM classification (log barrier method)

$$\min_{\substack{w,b,\\ \xi_i \in \mathbb{R}_+}} \Bigg[\|w\|^2 + C \sum_i^n \xi_i \\ - \frac{1}{t} \sum_i \left(log(-1 + y_i(\langle w, x_i \rangle + b) + \xi_i) + log(\xi_i) \right) \Bigg] \tag{11}$$

where C is a reqularization/penalty parameter

Optimization problem for finding the feasible starting point remains the same. Only the minimum value of s turns out to be negative for linearly inseparable dataset. This can be handled by appropriately chosing the initial values of $\xi_i \forall i$.

Result

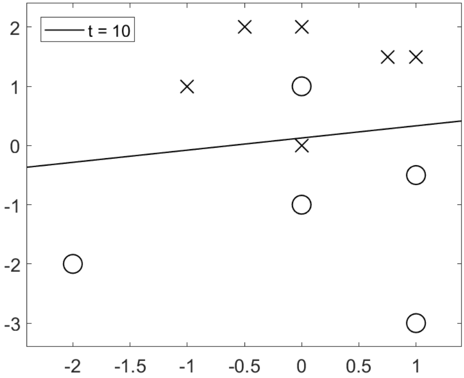

Figure 11: Separating hyperplane for a linearly inseparable data using log barrier method (t = 10; C = 1)

Result

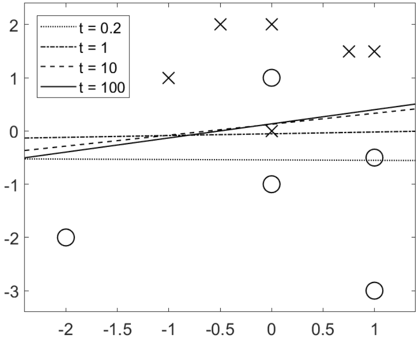

Figure 12: Separating hyperplane for a linearly inseparable data using log barrier method ($t = 0.2, 1, 10, 100$; $C = 1$) by incorporating the concept of central path [**Remark:** Central path method ultimately converges to exact method solution if the step size for t is adequate]

Comparison

Exact method

$$\begin{bmatrix}-0.2222 & 0.8889\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix}-0.1111=0$$

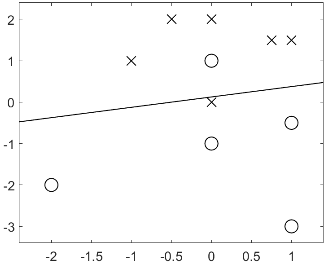

Figure 13: SVM using exact method

Log barrier method

$$\begin{bmatrix}-0.2364 & 0.8889\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix}-0.1199=0$$

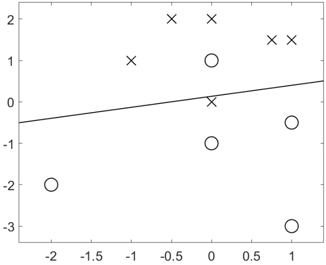

Figure 14: SVM using log barrier method and central path (t = 100)

anuragg.in@gmail.com

Scope

- Kernel operators
- Dual optimization problem

References

- S. Boyd and L. Vandenberghe.
  Convex optimization
  Cambridge University Press, 2004.
- A. Zisserman
  Lectures on Machine Learning, 2015.
  http://www.robots.ox.ac.uk/~az/lectures/ml/lect2.pdf.