Week 2 Report

Haoxuan Yin
Department of Computer Science and Technology
Tsinghua University
yhx18@mails.tsinghua.edu.cn

Abstract

Support Vector Machine is used to classify.It finds a best line to separate two sets.When it's difficult to do so,we can allow small errors or map the vectors to higher dimensions.

1 Original Problem

Given two sets of points ,we want to find a line to classify them i.e. to draw a line between the two sets whose distance to the nearest point is as far as possible.Mathematically,we want a function $f(x) = sgn(\omega^T x + b)$which returns 1 for one set and -1 for the other.In addition,we want to maximize γ , $s.t.\ \frac{1}{\|\omega\|} y^{(i)}(\omega^T x^{(i)} + b) \geq \gamma, i = 1, ..., m.$

After transformation,we equivalently need to minimize $\frac{1}{2}\|\omega\|^2$, $s.t.\ y^{(i)}(\omega^T x^{(i)} + b) \geq 1, i = 1, ..., m$.Then we only neeed to find a mathematician to solve this Lagrangian problem.

2 Improvements

There are times when the two sets are not linearly separable.

2.1 Allowing Errors

If the two sets intersect a bit,we can allow an error ξ_i and define a punishment constant C.Then the problem becomes to minimize $\frac{1}{2}\|\omega\| + C\sum_i \xi_i$, $s.t.\ y^{(i)}(\omega^T x^{(i)} + b) \geq 1 - \xi_i$

2.2 Kernel Function

With a proper function,we can transform x to a higher dimension so that the classification becomes easy.That is,we substitute all the x with $\Phi(x)$ where Φ is a mapping function.

Moreover,in the original problem,the ω we want can actuaaly be expressed in the form of $w = \sum_i \alpha_i y_i x_i$ and we only need to find out the best α_i.Notice that $\omega^T x + b = \sigma \alpha_i y^{(i)} \left\langle x^{(i)}, x \right\rangle + b$,so we only need to redefine the inner product.In fact,x_i doesn't even have to be a vector as long as you can define a positive-definite inner product.

3 Training the SVM

Firstly we define the kernel function $K(x, y)$ and punishment constant C.Then we can use a QP solver to directly calculate the α_i we want,but this approach is too slow.Instead,we introduce SMO(sequential minimal optimization) algrithm.We repeat until converge randomly choosing a pair α_i and α_j and optimize the target function while holding the other αs.