Example of Convex Optimization Problem

Support Vector Machine

(Konstantin Burlachenko, burlachenkok@gmail.com)

Problem name

Unconstrained Binary classification via support vector machine

Notes and references

Notes: $\prec$ elementwise, or the same as said that it is inequality in non-negative orthant cone.

Problem can be casted to this class: Quadratic programming, constraints are linear and objective is quadratic.

Interpretation

1) Seek affine function to approximately classify points.
2) Minimizing $|a|_2$ it the same as maximizing $\frac{2}{|a|_2}$ which is the width of the slab $1 \le a^T x - b \le 1$
3) Minimize misclassified points via L1 heuristic. And here in fact if you have enough computation power and number of variables (called features in machine learning) is not big (numbers like 15) then it's possible to fix misclassification pattern and solve sequence of *Convex Optimization* problems.
4) During solving this sequence of convex optimization problems duality can be used to early-termination of one of problems.

I think (1) and (2) is known to Machine Learning community, and (3),(4) is in fact mostly known for people who attack Machine Learning models from point of view of Math Optimization.

Problem formulation with explicit bias term

Objective to minimize	$\frac{\|a\|_2^2}{2} + \gamma(1^T u + 1^T v)$
Variables	$a \in R^n, b \in R, u \in R^{m_+}, v \in R^{m_-}$ m_+ - *number of positive examples* m_- - *number of negative examples*
Parameters	$\gamma \in R, x_i \in R^n - positive\ example, y_i \in R^n - negativ$
Explicit inequality constraint	$a^T x_i - b \geq 1 - u_i$ $a^T y_i - b \leq -(1 - v_i)$ $u_i \geq 0$ $v_i \geq 0$
Explicit equality constraint	$-$

Use predictor for Inference

$h(u) = a^T u - b$

If $a^T u - b \geq 0$ => we predict class “1”

If $a^T u - b \leq 0$ => we predict class “0”

Quadratic Optimization problem formulation with hidden bias term

Now assume for simplicity that constant factor for affine predictor is lie in one of the parameter due to that one feature is just constant “1”

Objective to minimize	$\frac{\|a\|_2^2}{2} + \gamma(1^T u + 1^T v)$
Variables	$a \in R^n, u \in R^{m_+}, v \in R^{m_-}$ m_+ - *number of positive examples* m_- - *number of negative examples*
Parameters	$\gamma \in R, x_i \in R^n - positive\ example, y_i \in R^n - negativ$
Explicit inequality constraint	$-a^T x_i \leq -1 + u_i$ $a^T y_i \leq -1 + v_i$ $-u_i \leq 0$ $-v_i \leq 0$

Objective to minimize	$\frac{\|a\|_2^2}{2} + \gamma(1^T u + 1^T v)$

Variables	$a\epsilon R^{n}, u\in R^{m_{+}}, v\in R^{m_{-}}$ m_{+} - *number of positive examples* m_{-} - *number of negative examples*
Parameters	$\gamma\epsilon R, x_{i}\epsilon R^{n} - positive\ example,$ $y_{i}\epsilon R^{n} - negative\ example$
Explicit inequality constraint	$-a^{T}x_{i} - u_{i} + 0\leq -1$ $+a^{T}y_{i} + 0 - v_{i}\leq -1$ $0 - u_{i} + 0\leq 0$ $0 + 0 - v_{i}\leq 0$

SVM problem formulation derivation with Hinge Loss

Here I will derive the reformulation which allow convert SVM problem as QP program to unconstrained minimization with *Hinge* Loss.

One SVM formulation is the following:

Objective: min. $\frac{|a|_{2}^{2}}{2} + \gamma(1^{T}u + 1^{T}v)$

Constraints:

$a^{T}x_{i} - b\geq 1 - u_{i}$

$a^{T}y_{i} - b\leq -(1 - v_{i})$

$u_{i}\geq 0$

$v_{i}\geq 0$

This optimization problem is the same as:

Objective: min. $\frac{|a|_{2}^{2}}{2} + \gamma(\sum max(u_{i}, 0) + \sum max(v_{i}, 0))$

Constraints:

$u_{i}\geq(1 - a^{T}x_{i} - b)(+1)$

$1 + (a^{T}y_{i} - b)\leq(v_{i})$

$u_{i}\geq 0$

$v_{i}\geq 0$

This problem is the same as:

Objective: $min. \frac{|a|_2^2}{2} + \gamma(\sum max(u_i, 0) + \sum max(v_i, 0))$

Constraints:

$u_i \geq (1 - a^T x_i - b)(s_i)$ *[1]*

$v_i \geq 1 - (a^T y_i - b)(s_i)$ *[2]*

$u_i \geq 0$ [3]

$v_i \geq 0$ [4]

Here I used extra notation trick. I used to $s_i = +1$ for positive examples and $s_i = -1$ for negative examples.

Before moving forward we can observe deep fact. Assume we found variable "a,b". Now making u_i or v_i greater then right hand side of inequalities [1] and [2] is in fact not optimal **if** right hand side is greater then zero.

But if right hand side of some constraints [1] or [2] is less then zero then due to constraints about non-negativity we should take u_i or v_i equal to zero.
Now if combine this two consideration then in fact we can combine everything into one objective and eliminate extra u_i and v_i completely.

Finally we can come to optimization problem formulation for SVM:

Objective: $min. \frac{|a|_2^2}{2} + \gamma(\sum max((1 - a^T x_i - b)(s_i), 0) + \sum max(1 - (a^T y_i - b)(s_i), 0))$

Notice about another notation:

Also sometimes people use other notation.

x_i – are both negative and positive examples to fit.

y_i – are sign of this examples.

Objective: $min. \frac{|a|_2^2}{2} + \gamma(\sum max((1 - a^T x_i - b)(y_i), 0))$

In both of this formulation objective is convex in (a,b), but not differentiable.