Coordinate descent Algorithm :

Let us say we want to optimize function $f(\alpha_1, \alpha_2, \dots \alpha_m)$

α_i's where $i \in \{1, 2, \dots, m\}$ are independent Variables.

Assume: No constraints on α_i's.

Algorithm:

Repeat until convergence {

FOR $i=1$ to m

$\alpha_i := \arg\min_{\hat{\alpha}_i} f(\alpha_1, \alpha_2 \dots \hat{\alpha}_i \dots \alpha_m)$ not fixed Rest of all assumed fixed

}

Problems: Coordinate descent algo for linear regression.

Let $f(x) = \frac{1}{2} \|Ax - y\|^2$ be the cost function to minimize,

where $y \in \mathbb{R}^n$, $A \in \mathbb{R}^{n \times p}$ with the columns $A_1, \dots A_p$

$x \in \mathbb{R}^p$

Consider minimizing over x_i, with all $x_j, j \neq i$ fixed

$$0 = \nabla_i f(x) = A_i^T (Ax - y) = A_i^T (A_i x_i + A_{-i} x_{-i} - y)$$

All except i

$$A_i^T A_i x_i + A_i^T A_{-i} x_{-i} = A_i^T y$$

$$\boxed{x_i = \frac{A_i^T (y - A_{-i} x_{-i})}{A_i^T A_i}}$$

Algorithm is as follows:

Repeat until convergence {

For $i = 1$ ~~to~~ p

$$x_i := \frac{A_i^T (y - A_{-i} X_{-i})}{A_i^T A_i}$$

}

<u>time Complexicity Analysis:</u>

$O(n)$ to compute $(y - A_{-i} X_{-i})$

$O(n)$ to compute $A_i^T (y - A_{-i} X_{-i})$

One cycle requires $O(np)$ operations.

Date: 29-06-2018

SMO (sequentially minimal optimization)

How to Apply coordinate descent Algorithm to constraint optimization Problems?

In svm Dual Problem:

$$\theta_D(\alpha) = \min_{\omega, b} \mathcal{L}(\omega, b, \alpha)$$

$$\max_{\alpha} \left(\theta_D(\alpha) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} \alpha_i \alpha_j y^{(i)} y^{(j)} \langle x^{(i)}, x^{(j)} \rangle\right)$$

$\alpha_i \geq 0 \quad \forall i \in \{1, 2, \ldots, m\}$

$\sum_{i=1}^{m} \alpha_i y^{(i)} = 0$, ← constraint

We can not change only one α_i (Keeping rest all fixed) because $\sum_{i=1}^{m} \alpha_i y^{(i)} = 0$.

Solution: instead of changing 1 α at a time. Change 2 α's at a time. This is called SMO - Sequential minimal optimization
(2 α's) (2 Minimum α's)

Algorithm:

Select α_i, α_j (how heuristics)

Hold all α_k's fixed except $k \neq i, j$

Optimize $\theta_D(\alpha)$ wrt to α_i, α_j Subject to all constraints

Constraint: $\sum_{i=0}^{m} \alpha_i y^{(i)} = 0$

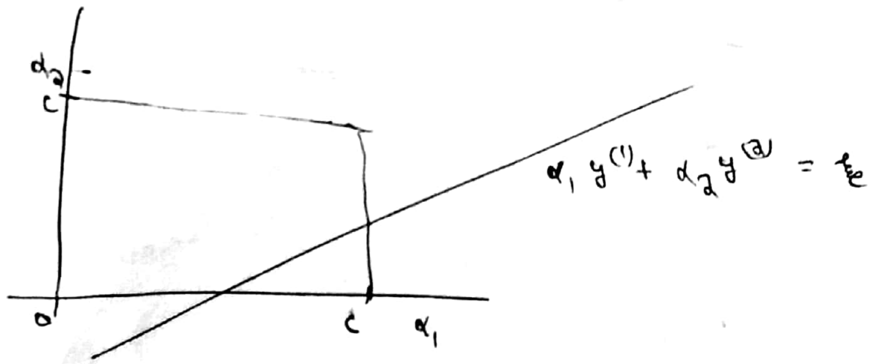
Let us change $\theta_D(\alpha)$ w.r.t α_1, α_2

$$\alpha_1 y^{(1)} + \alpha_2 y^{(2)} + \dots + \alpha_m y^{(m)} = 0$$

$$\alpha_1 y^{(1)} + \alpha_2 y^{(2)} = -\sum_{i=3}^{m} \alpha_i y^{(i)} = \zeta \quad \text{--- (1)}$$

For L1 norm Soft margin SVM:

$$0 \le \alpha_i \le C$$



from eqn. (1)

$$\alpha_1 = \frac{\zeta - \alpha_2 y^{(2)}}{y^{(1)}}$$

So, we can express $\theta_D(\alpha)$ as

$$\theta_D\left(\frac{\zeta - \alpha_2 y^{(2)}}{y^{(1)}}, \alpha_2, \alpha_3, \dots, \alpha_m\right)$$

θ_D is a function of only α_2 rest all are fixed

$$\theta_D = \sum_{i=1}^{m} \alpha_i - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} \alpha_i \alpha_j y^{(i)} y^{(j)} \langle x^{(i)}, x^{(j)} \rangle$$

↳ this will become similar to $ax^2 + bx + c = 0$

↳ optime this function.

Scanned by CamScanner