Sat Jan 1, 2005 7:24 AM

$$\max_{(x,y)\in D} \frac{1}{\|w\|}$$

s.t. $\forall (x,y)\in D \quad y(w^T\phi(x)+w_0)\geq 1$

Say the closest pt after solving does not have

$$y(w^T\phi(x)+w_0)=1$$

instead

$$y(w^T\phi(x)+w_0)>1 \quad \forall (x,y)\in D$$

$\Rightarrow$ I could st $w_{new} = \dfrac{w_{old}}{\min_{(x,y)\in D} y(w_{old}^T\phi(x)+w_{0\,old})}$

α

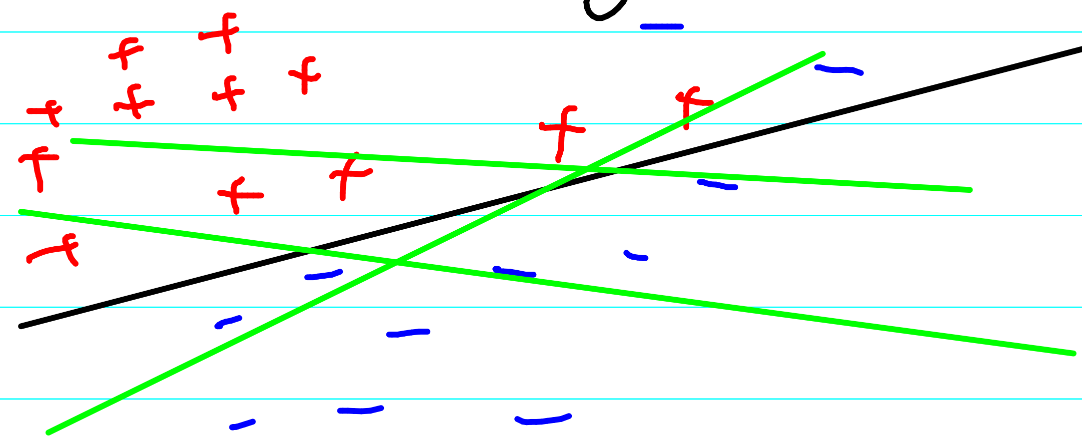
$$w_{0\,new} = \frac{w_{0\,old}}{\alpha}$$

s.t. $\min_{(x,y)\in D} y(w_{new}^T\phi(x)+w_{0\,new})=1$

and $\dfrac{1}{\|w_{new}\|} > \dfrac{1}{\|w_{old}\|}$

CONTRADICTION

Q: Can your hard margin SVM find a soln for this data?



Ans: No

Solution (in Black line): Find a classifier with maximum margin that has minimum misclassification

Contradictory. Therefore you need to tradeoff

Soft margin SVM

Should be monotonically increasing with ξ_{x_i}

could be $\xi_{x_i}^2$

$$\max_{(x,y)\in D} \frac{1}{\|w\|} - C\sum_{x\in D} \xi_x$$

$$\text{s.t. } \forall (x,y)\in D \quad y(w^T\phi(x)+w_0) \geq 1-\xi_x$$

$$\xi_x \geq 0$$

In practice, the most common soft margin SVM formulation is:

$\mathcal{D} = (\langle x_1, y_1\rangle, \langle x_2, y_2\rangle \dots \langle x_n, y_n\rangle)$

$$\min_{w, w_0} \; \|w\|^2 + C \sum_{i=1}^{n} \xi_i$$

could be replaced by ξ_i^2

$$\text{s.t. } i \in [1, \dots n] \quad y_i(w^T \phi(x_i) + w_0) \geq 1 - \xi_i$$

$$\text{s.t. } \forall i, \xi_i \geq 0$$

Q: If you use ξ_i^2 instead of ξ_i, the objective will shout if $\xi_i < 0$ even though first set of constraints are happy with $\xi_i < 0$

$\Rightarrow$ ξ_i will not be allowed to go < 0

$\therefore$ No need of constraint "$\xi_i \geq 0$"

Q: How to solve this opt prob

Consider optimizing

$\min x^2 + y^2$

Let $f(x, y) = x^2 + y^2$

$\min f(x, y)$ in this case,

consider soln to $\nabla f(x, y) = 0$

$$\nabla f(x_1 \ldots x_n) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}$$

Captures direction in which function f increases most rapidly.

[see pages 18-20 of optimization notes for connection between gradient & directional derivative & for proof]

Claim: If you are solving

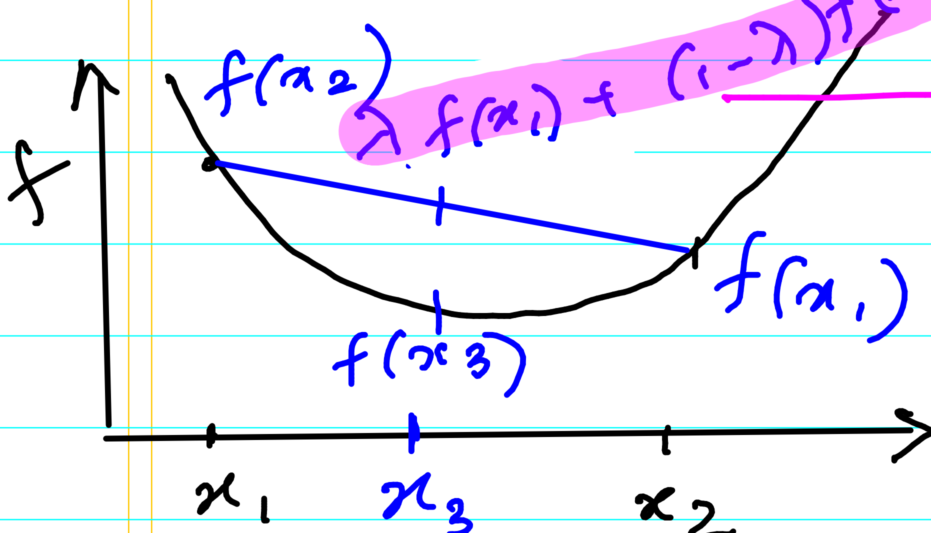
$$\min_{x \in R^n} f(x)$$

& f is continuous & differentiable in R^n then

If $x^* = \arg\min_{x \in R^n} f(x)$ (1)

then $\nabla f(x^*) = 0$ → Necessary condition for optimality

$\lambda f(x_1) + (1-\lambda) f(x_2)$ → convex combination of fn values



Convexity:

$$f(x_3) \leq \lambda f(x_1) + (1-\lambda) f(x_2)$$

$$x_3 = \lambda x_1 + (1-\lambda) x_2 \quad (\lambda \in [0,1])$$

Convex combination of the two pts x_1 & x_2

The condition

$$f(\lambda x_1 + (1-\lambda)x_2) < \lambda f(x_1) + (1-\lambda) f(x_2)$$

is called strict convexity

In general $x_1, x_2, x_3 \in \mathbb{R}^n$

Coming back to the optimization problem

if f is (strictly) convex & cts & differentiable

if $\nabla f(x^*) = 0$ then

$$x^* = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} f(x)$$

Sufficient condition for optimality

(2)

Q: How to characterize a (strictly) convex function more succintly and practically

Ans: Assuming $f(x)$

$$\left(x = [x_1 . x_2 \; . . . \; x_n]^T \in \mathbb{R}^n \right)$$

has continuous mixed partial derivatives

$$\begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & & & \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix} = \nabla^2 f$$

Hessian matrix

We can claim the following: abt the Hessian matrix :-

1) $\nabla^2 f$ is symmetric
when evaluated at any $x \in \mathbb{R}^n$

2) $\nabla f(x^*) = 0$ &

$\nabla^2 f(x^*)$ is positive definite

i.e $\forall v \in \mathbb{R}^n$ $v^T \nabla^2 f(x^*) v > 0$
& $v \neq 0$

is a necessary condition for x^* to be a local minimum for function f

3) If $\nabla^2 f(x)$ is positive definite $\forall x \in \mathbb{R}^n$, then f is strictly convex

Eg: $x^2 + y^2 = f(x, y)$ ③

$$\nabla^2 f(x,y) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \longrightarrow \begin{gathered} \forall v \neq 0 \\ v^T \nabla^2 f\, v \\ = 2v_1^2 + 2v_2^2 > 0 \end{gathered}$$

$\downarrow$

Another way of certifying positive definiteness (equivalent way)

All eigenvalues of $\nabla^2 f(x,y)$ should be > 0

To get eigenvalue λ

$$\det(\nabla^2 f - \lambda I) = 0$$

ie $\begin{vmatrix} 2-\lambda & 0 \\ 0 & 2-\lambda \end{vmatrix} = 0$

$$\Rightarrow \lambda_1 = 2, \ \lambda_2 = 2$$

Combining 2 & 3 & 1

if f is cts, differentiable & doubly differentiable & double derivatives are cts (condition could be slightly relaxed)

THEN

$$x^* = \arg\min_{x \in \mathbb{R}^n} f(x)$$ & f is strictly convex

Iff

$\nabla f(x^*) = 0$ & $\nabla^2 f(x)$ is positive definite $\forall x \in \mathbb{R}^n$

Some relaxations can be obtained by:

1) looking at positive semi-definiteness

ie $\forall v \neq 0$ $v^T \nabla^2 f(x) v \geq 0$

& this is equivalent to relaxing f to be convex

2) You can look at relaxed necessary condition

i.e

if $$x^* = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} f(x)$$

then $\nabla f(x^*) = 0$

& $\nabla^2 f(x^*)$ is positive (semi) definite

H/w: Refer to previous class notes & based on references provided there at the end, familiarize yourself with:

(a) lagrange multipliers

(b) necessary optimality conditions for constrained optimization

(c) necessary & sufficient conditions : Karush Kuhn Tucker conditions

(d) Duality theory & the dual optimization problem