5.4 2 category linearly separable case

slide1

sample : $y_1, .. y_n$ in $\omega_1 + \omega_2$

goal: find $\mathbf{a}$ for $g(x) = \mathbf{a}^t \mathbf{y}$

correct classification $\mathbf{a}^t y_i > 0$ if $y_i \in \omega_1$
< 0 if $y_i \in \omega_2$

slide2

normalization: drop labels $\omega_1 + \omega_2$
instead change sign of points in ω_2 s.t. $\mathbf{a}^t y_i > 0 \; \forall i$
i.e. all points positive

slide3

note: soln vector $\mathbf{a}$ is not unique

restriction:

soln vector must lie in intersection of
n half-spaces

What does this mean? Explain

concept of margin: Why might we want a margin

slide 4:

Key idea: $J(\mathbf{a})$ criterion fcn to be minimized

note: error in slide ⇒ maximization (is) / minimization (should be)

Learning rate $\eta(k)$ — BIG ISSUE
How to set

slide 5 Principled method for setting η

First expression for $J(a)$
↑ soln vector

second expression: replace $a - a(k)$ in 1st expression with ∇J
scale by η

Third expression: choice of η (optimal choice)

Newton's alg. simpler update rule
→ still need to invert H.

slide 6 comparison of convergence for simple gradient & Newton's Alg.

slide 7 Perceptron criterion function J_p

idea: try to find cheaper than $J()$

1st bad idea: $J() = \#$ of misclassified examples

skip to next slide to show graph

better idea: J_p — error in distance
↑
sum

note: ∇J_p is very simple: $\sum_{y \in \mathcal{Y}} -y$

update is simple

note: still have issue of selecting η

slide 8 contrast $J(a)$ + $J_p(a)$

slide 9 Explain figure

1) note a is normal to the hyperplane boundary

$\Rightarrow \forall i \; \angle a y_i < 90°$

Thus left boundary of soln region is 90° wrt y_3
right " wrt y_1

i.e. intersection of n half-spaces

$\Rightarrow$ update: adding $y_i \in \mathcal{Y}$ displaces a as shown

5.4.4

~~slide~~ slide 10

repeatedly cycle through y until all correctly classified

sequence: misclassified

$y_1\ y_2\ y_3\ y_1\ y_2\ y_3\ y_1\ y_2 \ldots$

violation: $y^k \leftarrow k^{th}$ misclassified sample

Alg 4. instead of batch $\Rightarrow$ 1 correction per iteration

slide 11 note: introduction of scaling factor α update reduces distance to soln $\hat{a}$ ~~by y~~

note: squared distance ~~shows~~ illustrates distance reduction

slide 12 obs: $a^t(k) y^k \leq 0$ since y^k misclassified

picking large α then $2\alpha \hat{a}^t y^k$ dominates $\|y^k\|^2$

$\Rightarrow$ HOW to pick α?

let $\beta^2 = \max_i \|y_i\|^2$

let $\gamma = \min_i [\hat{a}^t y_i] > 0$

choose $\alpha = \frac{\beta^2}{\gamma}$

Have to know soln $\hat{a}$ to specify γ

5.4.5

slide 13 obs update reduces distance by β^2

↑ at least

or i.e $\max_i \|y_i\|^2$

eq. 25 specifies bound on # of corrections

slide 14 note: we need a soln $\hat{a}$ to compute the bound!

bound in terms of max vector + smallest soln distance from hyperplane

slide 15 Generalizations

1) variable learning rate $\eta(k)$

2) margin b