01/24/2017.

① A Tutorial on MMA (majorization minimization algorithm)

② An MMA for multi-category vertex discrinant Analysis.

min $f(\theta)$. use $g(\theta|\theta^{(m)})$ to approximate $f(\theta)$

convex function.

Step 1: start point $\theta^{(0)}$

2: Find $g(\theta|\theta^{(m)})$

(i) $g(\theta|\theta^{(m)}) \geq f(\theta)$

(ii) $g(\theta^{(m)}|\theta^{(m)}) = f(\theta^{(m)})$.

$\theta^{(m+1)} = \arg\min g(\theta|\theta^{(m)})$

3. Repeat 2 until converge.

Claim: $f^{(m+1)}(\theta) \leq f^{(m)}(\theta)$.

proof: $f^{(m+1)}(\theta) = g(\theta^{(m+1)}|\theta^{(m)}) + \underbrace{f(\theta^{(m+1)}) - g(\theta^{(m+1)}|\theta^{(m)})}_{\leq 0}$

since ~~$g(\theta^{m}$~~ $\theta^{(m+1)} = \arg\min g(\theta|\theta^{(m)})$

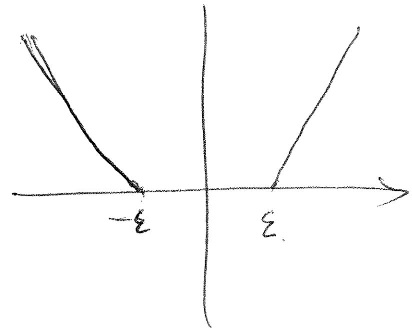
$\leq g(\theta^{(m)}|\theta^{(m)}) + f(\theta^{(m)}) - g(\theta^{(m)}|\theta^{(m)}) = f(\theta^{(m)})$.

K-class classification.

$$|y - \underset{p\times 1}{a'X} - b|_\varepsilon$$

ε - insentive Loss:

$$|z|_\varepsilon = \max\{|z| - \varepsilon, 0\}.$$

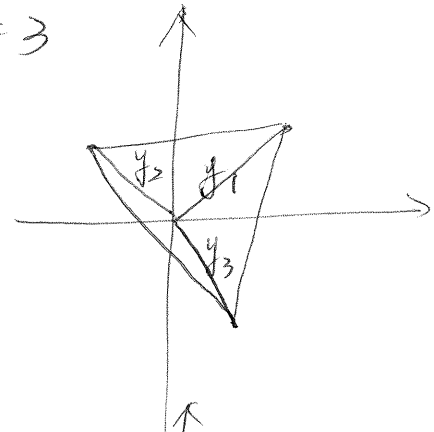


if $K = 2$ $\quad -1$

$\qquad\qquad\quad 1$

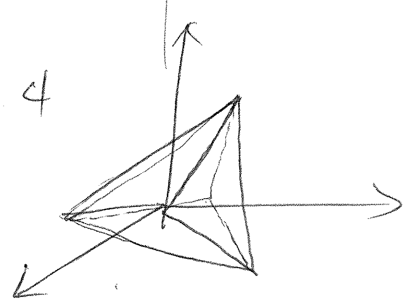
if $K > 2$, use $(K-1)\times 1$ vector to express each class.

K = 3



$$y_j = \begin{cases} K^{-\frac{1}{2}}\,\bar{I} & \text{if } j = 1 \\ -\dfrac{1+\sqrt{K+1}}{K^{3/2}}\cdot \bar{I} + \sqrt{\dfrac{K+1}{K}}\, e_{j-1}, & j > 1 \end{cases}$$

K = 4



extension $\quad \|z\|_\varepsilon = \max\{\|z\|_2 - \varepsilon, 0\}$

$$\|y - \underset{(K-1)\times p}{A^T} X - \underset{(K-1)\times 1}{b}\|_\varepsilon$$

$$\min_{A,b} E\|y - A^T X - b\|_\varepsilon$$

$$\min_{A,b} R(A,b) = \frac{1}{n}\sum_{i=1}^{n}\|y_i - AX_i - b\|_\varepsilon + \lambda\sum_{j=1}^{K-1}\|a_j\|^2 \qquad (1)$$

a_j' is jth row of A.

Convex problem.

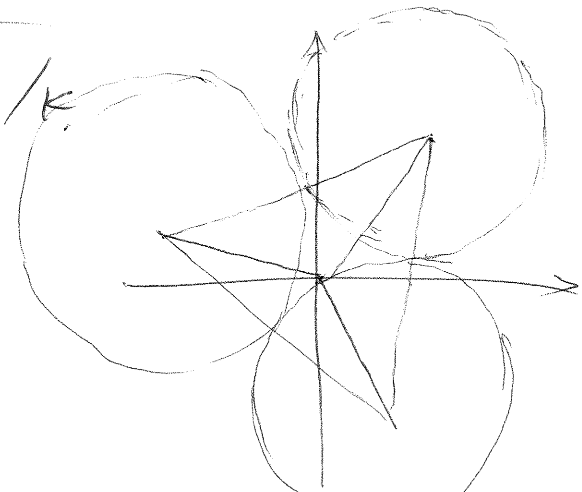
How to majorize first part?

Claim: if $f(x) = \|x\|_\varepsilon$,
$$g(x|x^{(m)}) = \begin{cases} \dfrac{1}{2\|x^{(m)}\|}\|x\|^2 + \dfrac{1}{2}\|x^{(m)}\| - \varepsilon, & \text{for } \|x^{(m)}\| \geq 2\varepsilon, \\ \dfrac{1}{4(\varepsilon - \|x^{(m)}\|)}\|x - x^{(m)}\|^2 & \text{for } \|x^{(m)}\| < \varepsilon \\ \dfrac{1}{4(\varepsilon - \|c(x^{(m)})\cdot x^{(m)}\|)}\|x - c\cdot x^{(m)}\|^2, & \text{for } \varepsilon \leq \|x^{(m)}\| < 2\varepsilon \end{cases}$$

where $c = \left(\dfrac{2\varepsilon}{\|x^{(m)}\|} - 1\right)$

$$(1) \leq \frac{1}{n}\sum_{i=1}^{n} w_i \|r_i - v_i\|^2 + \lambda \sum_{j=1}^{K-1}\|a_j\|^2 + f.$$
$$= \frac{1}{n}\sum_{i=1}^{n} w_i \sum_{j=1}^{K-1}\left((y_{ij} - v_{ij}) - a_j' x_{ij} - b_j\right)^2 + \lambda\sum_{j=1}^{K-1}\|a_j\|^2 + f.$$
$$= \sum_{j=1}^{K-1}\left[\frac{1}{n}\sum_{i=1}^{n} w_i (y_{ij} - v_{ij}) - a_j' x_j - b_j)^2 + \lambda\|a_j\|^2\right]$$

suggestion for $\varepsilon = \frac{1}{2}\sqrt{(2K+2)/K}$.



speed of convergence is linear.

$$\lim_{m\to\infty} \frac{\|\theta^{(m+1)} - \theta^*\|}{\|\theta^{(m)} - \theta^*\|} = c < 1$$