\[ g_i(x) = \max_{j=1,\ldots,n_i} g_{ij}(x) \]

\[ g_{ij}(x) = w_{ij}^T x + w_{ijo} \quad \bar{v} = 1, \ldots, c \]
\[ j = 1, \ldots, n_i \]

Decision regions are
\[ R_i^c = \left\{ x : g_i(x) > g_k(x) \neq k \neq i \right\} \]
\[ = \left\{ x : \max_j g_{ij}(x) > \max_l g_{kl}(x) \neq k \neq i \right\} \]

Define a linear machine whose decision regions are
\[ R_{ij} = \left\{ x : g_{ij}(x) > g_{kl}(x) \neq i \neq k \lor j \neq l \right\} \]

Then \[ R_i = \bigcup_{j=1}^{n_i} R_{ij} \]

\( R_{ij} \) is a linear machine, \( R_{ij} \) are convex w/ piece-wise hyperplane boundaries.

\[ R_i = \bigcup_{j=1}^{n_i} R_{ij} \] is a union of convex regions, which need not be connected or convex.

\[ g_{1i}(x) = -1 - x \]
\[ g_{2i}(x) = x - 1 \]
\[ g_{21}(x) = 0 \]
Define the $k^{th}$ correction to be

$$a(k+1) = a(k) + \eta(k) y(k)$$

$0 < \eta_a \leq \eta(k) \leq \eta_b < \infty$,

- $y(k)$ drawn uniformly (with replacement) from the set $\{ \bar{y}_i : \bar{a}(k)^T \bar{y}_i \leq b \}$

- Let $\hat{a}$ be any solution vector s.t. $\hat{a}^T y(k) \geq y > \emptyset \not\equiv k$

- Define $\beta^2 = \max_i \| y_i \|_2^2$

Then

$$\| a(k+1) - \alpha \hat{a} \|_2^2 = \| a(k) - \alpha \hat{a} \|_2^2 + 2\eta(k) a(k)^T y(k) - 2\alpha\eta(k) \hat{a}^T y(k) + \eta^2(k) \| y(k) \|_2^2$$

$$\leq \| a(k) - \alpha \hat{a} \|_2^2 + 2\eta \beta - 2\eta_b \alpha \gamma + \eta^2 \beta^2$$

If we choose

$$\alpha > \left( \frac{\eta_a}{\eta_b} \right) \left( \frac{\beta^2}{\gamma} \right)$$

If we choose $b \leq -\frac{1}{2\alpha} \eta \beta^2 = -\frac{\eta_a \max_i \| y_i \|_2^2}{2 \min \hat{a}^T y_i}$

then monotonous reduction of error is guaranteed from the very first step ($\alpha = 0$)!
Write \( \hat{a}(k+1) = a(k) - \eta(k) Y^T \tilde{s}(k) \).

For batch update, \( \tilde{s}(k) = \tilde{e}(k) = Y \hat{a}(k) - b \).

For single-sample update, \( \tilde{s}(k) = Y_{i(k)}^T \hat{a}(k) - b_{i(k)} \), where \( i(k) \) is sample chosen at time \( k \).

\[ \eta(k) = \frac{\eta}{k}, \quad \eta \equiv \eta(1) \]

Demonstrate that \( \hat{a}(k) \rightarrow \hat{a} \)

\[ s.t. \quad Y^T (Y \hat{a} - b) = 0 \]

\[ Y^T = (Y^TY)^{-1} Y^T \]

\[ \Leftrightarrow \quad Y^T(Y \hat{a} - b) = 0 \quad \text{assuming } Y^TY \text{ invertible} \]

\[ \| \hat{a}(k+1) - \hat{a} \|_2^2 = \| \hat{a}(k) - \hat{a} \|_2^2 - 2 \left( \frac{\eta}{k} \right) \tilde{s}(k)^T Y \hat{a}(k) \]

\[ + 2 \left( \frac{\eta}{k} \right) \tilde{s}(k)^T Y \hat{a} + \left( \frac{\eta}{k} \right)^2 \| Y^T \tilde{s}(k) \|_2^2 \]

\[ = \| \hat{a}(k) - \hat{a} \|_2^2 - 2 \left( \frac{\eta}{k} \right) \tilde{s}(k)^T (\tilde{e}(k) + b) \]

\[ + 2 \left( \frac{\eta}{k} \right) \tilde{s}(k)^T Y \hat{a} + \left( \frac{\eta}{k} \right)^2 \| Y^T \tilde{s}(k) \|_2^2 \]

\[ = \| \hat{a}(k) - \hat{a} \|_2^2 + 2 \left( \frac{\eta}{k} \right) \tilde{s}(k)^T [\tilde{e} - \tilde{e}(k)] + \left( \frac{\eta}{k} \right)^2 \| Y^T \tilde{s}(k) \|_2^2 \]

\[ \leq \| \hat{a}(k) - \hat{a} \|_2^2 + 2 \left( \frac{\eta}{k} \right) \| \tilde{s}(k) \|_2 \| \tilde{e} \|_2 - 2 \left( \frac{\eta}{k} \right) \| \tilde{s}(k) \|_2 \max_{1 \leq i \leq n} \| \tilde{e}_i \|_2 \]

\[ + \left( \frac{\eta}{k} \right)^2 \| \tilde{s}(k) \|_2^2 \]

\[ \leq \| \hat{a}(k) - \hat{a} \|_2^2 \]

for all \( k \) s.t.

\[ \tilde{s} \equiv \hat{a} = Y \hat{a} - b \quad \text{in batch mode} \]

\[ \tilde{s} = Y_{i(k)}^T \hat{a} - b_{i(k)} \quad \text{in single-sample mode} \]

so it converges as long as \( \| \tilde{s} \|_2 < \| \tilde{s}(k) \|_2 \) on average.
5.33

(a) \[ L(\alpha, \ddot{\alpha}) = -\frac{1}{2} \| \ddot{\alpha} \|_2^2 - \sum_{k=1}^{\hat{\alpha}} \alpha_k [Z_k \ddot{\alpha}^T \ddot{y}_k - 1] \]

The desired solution is a saddle point:
\[ \alpha^*, \ddot{\alpha}^* = \arg\min_{\alpha} \arg\max_{\ddot{\alpha}} L(\alpha, \ddot{\alpha}) \]

because this Lagrangian was constructed from the original problem:
\[ \text{minimize } \frac{1}{2} \| \ddot{\alpha} \|_2^2 \text{ s.t. } \alpha_k = 0 \text{ (no error)} \]
\[ \text{or } Z_k \ddot{\alpha}^T \ddot{y}_k < 1 \text{ (error)} \]

\( \alpha_k \) should be as large as is required to ensure fulfillment of the constraint, but no larger; larger \( \alpha_k \) would unnecessarily increase \( \frac{1}{2} \| \ddot{\alpha} \|_2^2 \)

(b) This step requires a notational oddity covered in lecture but omitted in the text. We assume
\[ \ddot{\alpha}^T \ddot{y}_k = a_0 + a_1 y_{1k} + \cdots + a_d y_{dk} \]
but \[ \| \ddot{\alpha} \|_2^2 = a_1^2 + a_2^2 + \cdots + a_d^2 \]

\[ \frac{\partial L}{\partial a_0} = -\sum_{k=1}^{\hat{\alpha}} \alpha_k Z_k = 0 \text{ at } \alpha_k = \alpha_k^* \]

\[ \frac{\partial L}{\partial a_i} = a_i - \sum_{k=1}^{\hat{\alpha}} \alpha_k Z_k y_{ik}, \quad 1 \leq i \leq d \]

setting \[ \frac{\partial L}{\partial a_i} = 0 \] we get \[ \ddot{\alpha}^* = \sum_{k=1}^{\hat{\alpha}} \alpha_k Z_k \ddot{y}_k \]

(c) \[ \frac{\partial L}{\partial \ddot{\alpha}_i} = \ddot{\alpha}_i - \sum_{k=1}^{\hat{\alpha}} \alpha_k Z_k \ddot{y}_k, \quad 1 \leq i \leq d \]

setting \[ \frac{\partial L}{\partial \ddot{\alpha}_i} = 0 \] we get \[ \ddot{\alpha}^* = \sum_{k=1}^{\hat{\alpha}} \alpha_k Z_k \ddot{y}_k \]

(d) \[ \alpha_k^* = 0 \] or \[ [Z_k \ddot{\alpha}^T \ddot{y}_k - 1] = 0 \] \[ \Rightarrow \alpha_k^* [Z_k \ddot{\alpha}^T \ddot{y}_k - 1] = 0 \text{ at } k \]
By the associative law of multiplication over addition,
\[ \frac{1}{2} ||a||^2 - \sum_{k=1}^{\hat{n}} \alpha_k \left[ z_k \hat{a}^T y_k - 1 \right] = \frac{1}{2} ||a||^2 - \sum_{k=1}^{\hat{n}} \alpha_k z_k \hat{a}^T y_k + \sum_{k=1}^{\hat{n}} \alpha_k \]

Substituting \( \hat{a} = \sum_{k=1}^{\hat{n}} \alpha_k z_k \hat{y}_k \):
\[ L(\hat{a}, a) = \frac{1}{2} \sum_{k=1}^{\hat{n}} \sum_{l=1}^{\hat{n}} \alpha_k \alpha_l z_k z_l \hat{y}_k \hat{y}_l - \sum_{k=1}^{\hat{n}} \sum_{l=1}^{\hat{n}} \alpha_k \alpha_l z_k z_l \hat{y}_k \hat{y}_l + \sum_{k=1}^{\hat{n}} \alpha_k \]
\[ = -\frac{1}{2} \sum_{k=1}^{\hat{n}} \alpha_k \alpha_l z_k z_l \hat{y}_k \hat{y}_l + \sum_{k=1}^{\hat{n}} \alpha_k \]