Multiple Discriminant Analysis

basic idea: generalize Fisher's linear discriminant to c-classes & $c-1$ discriminant functions

Q: We project from d dimensions to ?
A: $c-1$ dimensions. We assume $d \geq c$

note: we generalize the concept of within-class scatter matrix

$$S_W = \sum_{i=1}^{c} S_i \quad \text{i.e. sum of all inclass scatter}$$

recall $S_i = \sum_{x \in D_i} (x - m_i)(x - m_i)^t$

and $m_i = \frac{1}{n_i} \sum_{x \in D_i} x_i$

Q: How do we generalize between class scatter?
recall in FLD we had $S_B = (m_1 - m_2)(m_1 - m_2)^t$

Obs. No easy way to ~~???~~ extend to c means!

Plan B: 1) find Total scatter
2) subtract S_W leaving S_B

Total scatter $\quad S_T = \sum_{x} (x-m)(x-m)^t$

where $\quad m = \frac{1}{n} \sum_{x} x$

now we need to separate S_T into its S_W + S_B components

obs. $\quad (x - m_i + m_i - m)(x - m_i + m_i - m)^t = (x-m)(x-m)^t$

$$\Rightarrow \quad S_T = \sum_{i=1}^{c} \sum_{x \in D_i} (x - m_i + m_i - m)(x - m_i + m_i - m)^t$$

$$= \underbrace{\sum_{i=1}^{c} \sum_{x \in D_i} (x - m_i)(x - m_i)^t}_{S_W} + \underbrace{\sum_{i=1}^{c} \sum_{x \in D_i} (m_i - m)(m_i - m)^t}_{S_B}$$

We project onto $c-1$ space with $c-1$ discriminants

$$y_i = w_i^t x \quad , \quad i = 1, \ldots, c-1$$

↑ dimension i of projection

The vector of $c-1$ dimensions : $\quad y = W^t x$

where each w_i is a column in W

Now we need to show how S_B + S_W are projected into $c-1$ dimensions

Q: why do we need this?

A: we want to find W that best discriminates our c classes
we need to evaluate $S_B + S_W$ in $c-1$ dimensions

1) we can use our formulas for: eqs 119 – 122

m_i, m, S_W + S_B to find

$\tilde{m}_i$, $\tilde{m}$, $\tilde{S}_W$ + $\tilde{S}_B$ from $y = W^t x$

we can also directly project $S_W \rightarrow \tilde{S}_W$ + $S_B \rightarrow \tilde{S}_B$

i.e. $\tilde{S}_W = W^t S_W W$ + $\tilde{S}_B = W^t S_B W$

goal: find a criterion fcn $J(W)$ analogous to $J(w)$ for FLD

→ same idea: maximize ratio of between class scatter / within class scatter

$$\Rightarrow J(W) = \frac{|\tilde{S}_B|}{|\tilde{S}_W|}$$ i.e. we take ratio of the determinants

obs: the columns of an optimal W are the eigenvectors of the largest ~~eigenvalues~~ eigenvalues in

$$S_B w_i = \lambda_i S_W w_i$$

→ We could solve as a conventional eigenvalue problem

⇒ need to find inverse of S_W

instead solve $|S_B - \lambda_i S_W| = 0$ (doesn't depend on W) we get λ_i's

then solve $(S_B - \lambda_i S_W) w_i = 0$ to get w_i's (eigenvectors)

<u>note:</u> no more than $c-1$ eigenvalues are nonzero

(because S_B is the sum of c matrices of rank one or less + only $c-1$ are independent)

<u>Next:</u> use methods from chapt. 2 to create full classifier