Fisher's Linear Discriminant

- To maximise a fn: large separation b/w projected class means, & a small variance within each class
- max inter-class, min intra-class: one criterion $\Longrightarrow$
- To maximise inter/intra: $\boxed{J(\mathbf{w}) = \frac{(m_2 - m_1)^2}{s_1^2 + s_2^2}}$
- $s_j \stackrel{\triangle}{=} \sum_{i \in \mathscr{C}_j} (y_i - m_j)^2 / N_j, \qquad y_i = \mathbf{w}^T \mathbf{x}_i$
- $J(\mathbf{w})$ dim'less: means-diff-sq/variances-sum
- Can't normalise Type-1/Type-2 else N^r = 0, prob!
- Numerator = $(m_2 - m_1)^2 = \{\mathbf{w}^T(\mathbf{m}_2 - \mathbf{m}_1)\}^2 =$ scalar! $= \{\mathbf{w}^T(\mathbf{m}_2 - \mathbf{m}_1)\}\{\mathbf{w}^T(\mathbf{m}_2 - \mathbf{m}_1)\}^T = \mathbf{w}^T(\mathbf{m}_2 - \mathbf{m}_1)(\mathbf{m}_2 - \mathbf{m}_1)^T\mathbf{w} = \mathbf{w}^T \mathbf{S}_B \mathbf{w}$
  $\mathbf{S}_B$: Between-Class Covariance
- Denominator = $\sum s_j^2$. Now, $s_j^2 = \sum_{i \in \mathscr{C}_j} (y_i - m_j)^2 / N_j =$ scalar! $= \frac{1}{N_j}\{\mathbf{w}^T(\mathbf{x}_i - \mathbf{m}_j)\}\{\mathbf{w}^T(\mathbf{x}_i - \mathbf{m}_j)\}^T$

- Denominator = $\mathbf{w}^T [\frac{1}{N_1} \sum_{i \in \mathscr{C}_1} (\mathbf{x}_i - \mathbf{m}_1)(\mathbf{x}_i - \mathbf{m}_1)^T + \frac{1}{N_2} \sum_{i \in \mathscr{C}_2} (\mathbf{x}_i - \mathbf{m}_2)(\mathbf{x}_i - \mathbf{m}_2)^T] \mathbf{w} = \mathbf{w}^T \mathbf{S}_W \mathbf{w}$
- $\boxed{J(\mathbf{w}) = \frac{\mathbf{w}^T \mathbf{S}_B \mathbf{w}}{\mathbf{w}^T \mathbf{S}_W \mathbf{w}}}$
- $\frac{\partial J(\mathbf{x})}{\partial \mathbf{w}} = 0 \implies \frac{(\mathbf{w}^T \mathbf{S}_W \mathbf{w}) 2 \mathbf{S}_B \mathbf{w} - (\mathbf{w}^T \mathbf{S}_B \mathbf{w}) 2 \mathbf{S}_W \mathbf{w}}{(\mathbf{w}^T \mathbf{S}_W \mathbf{w})^2} = 0$
  $\mathbf{S}_B \& \mathbf{S}_W$: data-dep consts $\implies \mathbf{S}_W \mathbf{w} = (\frac{\mathbf{w}^T \mathbf{S}_W \mathbf{w}}{\mathbf{w}^T \mathbf{S}_B \mathbf{w}}) \mathbf{S}_B \mathbf{w}$
- $\implies \mathbf{w} = \frac{1}{J(\mathbf{w})} \mathbf{S}_W^{-1} (\mathbf{m}_2 - \mathbf{m}_1)(\mathbf{m}_2 - \mathbf{m}_1)^T \mathbf{w}$
- $= \frac{1}{J(\mathbf{w})} \mathbf{S}_W^{-1} (\mathbf{m}_2 - \mathbf{m}_1) \{\mathbf{w}^T (\mathbf{m}_2 - \mathbf{m}_1)\}^T$
- $\implies \mathbf{w} = \frac{m_2 - m_1}{J(\mathbf{w})} \mathbf{S}_W^{-1} (\mathbf{m}_2 - \mathbf{m}_1)$ $\boxed{\mathbf{w} \propto \mathbf{S}_W^{-1} (\mathbf{m}_2 - \mathbf{m}_1)}$
- Fisher's result: Weights depend on the difference in the means & the distribution/overall covariance
- Fisher: not a discriminant, but gives a direction for 1-D projection $\mathbf{w}$. $y(\mathbf{x}) = \mathbf{w}^T \mathbf{x} > / < Thresh$