Home Page

Title Page

Contents

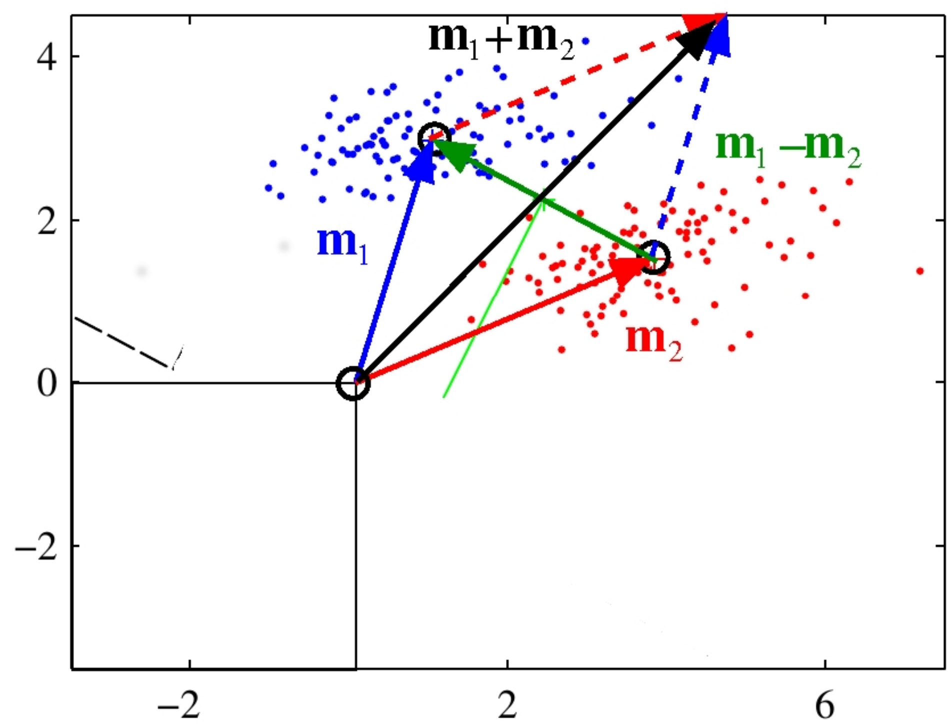
Page 25 of 30

Go Back

Full Screen

Close

Quit



- 2-D point $\mathbf{m}_j$ is a position vector: joining the origin to the point
- Triangle law of vectors: the line $\mathbf{z}$ joining $\mathbf{m}_2$ to $\mathbf{m}_2$ is $\mathbf{m}_1 - \mathbf{m}_2$. $(\mathbf{m}_2 + \mathbf{z} = \mathbf{m}_1)$
- Parallelogram law: main diag $\mathbf{m}_1 + \mathbf{m}_2$

Home Page

Title Page

Contents

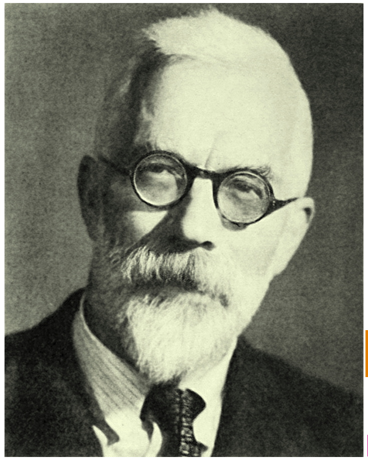
Page 26 of 30

Go Back

Full Screen

Close

Quit



Fisher's Linear Discriminant

R. A. Fisher [1890-1962]

https://upload.wikimedia.org/wikipedia/commons/3/37/Biologist_and_statistician_Ronald_Fisher.jpg

- Development: 2-class: $y(\mathbf{x}) = y(\mathbf{x}, \mathbf{w}) = \mathbf{w}^T\mathbf{x} + w_0$
- Call $\mathscr{C}_1$ if $y(\mathbf{x}) \geq 0$ i.e., $\mathbf{w}^T\mathbf{x} + w_0 \geq 0$, else call $\mathscr{C}_2$
- $\mathbf{w}^T\mathbf{x}$: projection of $D-$dim data onto 1-D
- Phy Sig of $\mathbf{w}^T\mathbf{x} > -w_0$: comparing with a thresh
- Comment: projecting onto 1-D may lead to considerable loss of info; classes well-separated in $D-$D may strongly overlap in 1-D (projection!)
- However: Adjusting components of $\mathbf{w}$: can select a projection that maximises the class separation

Start from a 2-class problem

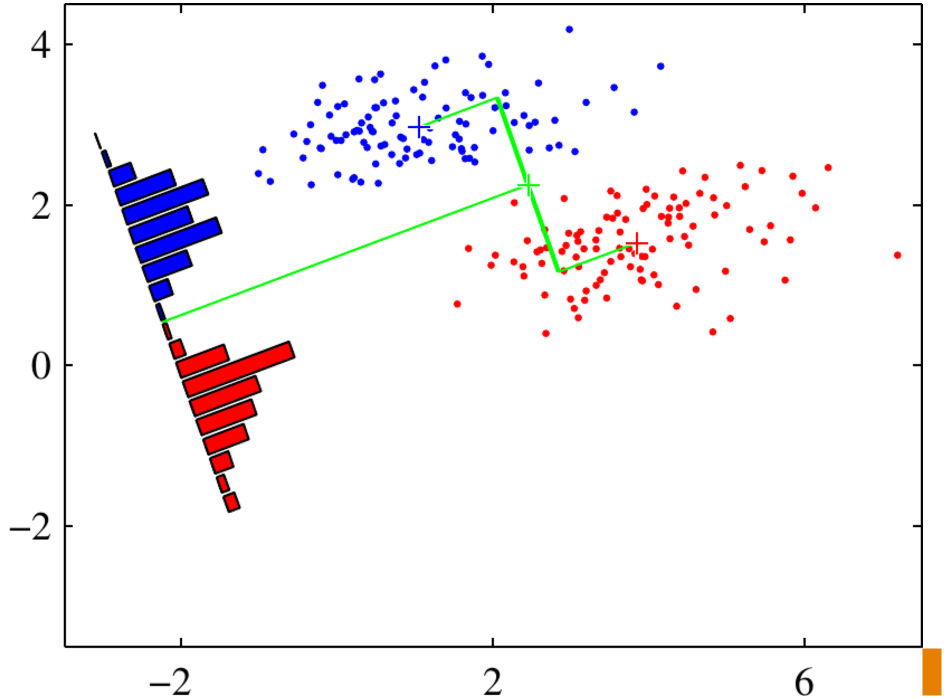
$\mathscr{C}_1 : N_1$ pts; $\mathbf{m}_1 = \frac{1}{N_1}\sum_{i \in \mathscr{C}_1} \mathbf{x}_i$; $\mathscr{C}_2 : N_2$ pts; $\mathbf{m}_2 = \frac{1}{N_2}\sum_{i \in \mathscr{C}_2} \mathbf{x}_i$

- Attempt 1: Simplest measure of class separation (when projected onto the $\mathbf{w}$): separation of the projected class means; $m_1 = \mathbf{w}^T\mathbf{m}_1; m_2 = \mathbf{w}^T\mathbf{m}_2$
- $\boxed{m_2 - m_1 = \mathbf{w}^T(\mathbf{m}_2 - \mathbf{m}_1)}$ Choose $\mathbf{w}$ to max $m_2 - m_1$

1. Problems! Can select $\mathbf{w}$ arbitrarily large
2. $\frac{\partial(m_1 - m_2)}{\partial \mathbf{w}} = 0 \implies m_2 - m_1 = 0$: Minimum!

- Attempt 2: Constrained Optimisation: Find the weight vector among the infinite with unit norm
- $f(\mathbf{w}) = \mathbf{w}^T(\mathbf{m}_2 - \mathbf{m}_1) + \lambda(\mathbf{w}^T\mathbf{w} - 1) = (\mathbf{m}_2 - \mathbf{m}_1)^T\mathbf{w} + \lambda(\mathbf{w}^T\mathbf{w} - 1)$. $\frac{\partial f(\mathbf{w})}{\partial \mathbf{w}} = 0 : (\mathbf{m}_2 - \mathbf{m}_1)^T + 2\lambda\mathbf{w}^T = 0 \implies \mathbf{w} = -2\lambda(\mathbf{m}_2 - \mathbf{m}_1) \propto (\mathbf{m}_2 - \mathbf{m}_1)$
- Problem: 2 classes well-separated in the original 2-D space may have considerable overlap when projected onto the line joining the means

[C. M. Bishop, Pattern Recognition and Machine Learning. Springer, 2006. Fig. 4.6, p. 188]



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: $\boxed{J(\mathbf{w}) = \frac{(m_2 - m_1)^2}{s_1^2 + s_2^2}}$
- $s_j \triangleq \sum_{i \in \mathscr{C}_j} (y_i - m_j)^2 / N_j, \quad y_i = \mathbf{w}^T \mathbf{x}_i$
- $J(\mathbf{w})$ 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! $= \frac{1}{N_j} \{\mathbf{w}^T(\mathbf{x}_i - \mathbf{m}_j)\}\{\mathbf{w}^T(\mathbf{x}_i - \mathbf{m}_j)\}^T$

Home Page
Title Page
Contents
Page 30 of 30
Go Back
Full Screen
Close
Quit

- 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 $\mathbf{w}$. $y(\mathbf{x}) = \mathbf{w}^T\mathbf{x} > / < Thresh$