Home Page

Title Page

Contents

Page 22 of 30

Go Back

Full Screen

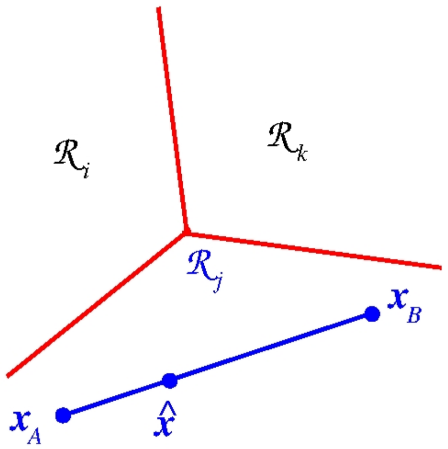
Close

Quit

Elegant: $K-$ Class Discriminant!

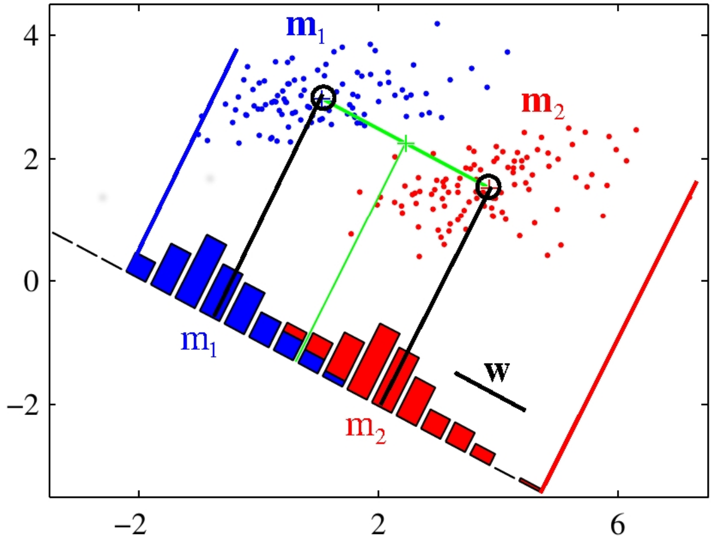
$$y_j(\mathbf{x}) = \mathbf{w}_j^T\mathbf{x} + w_{j0}$$

- Decision rule: Class $\mathscr{C}_j$ if $y_j(\mathbf{x}) > y_l(\mathbf{x}) \forall j \neq j$
- Decision boundary b/w $\mathscr{C}_j$ & $\mathscr{C}_k$: $y_j(\mathbf{x}) = y_k(\mathbf{x})$
- $\implies \mathbf{w}_j^T\mathbf{x} + w_{j0} = \mathbf{w}_k^T\mathbf{x} + w_{k0}$
- $\implies (\mathbf{w}_j - \mathbf{w}_k)^T\mathbf{x} + (w_{j0} - w_{k0}) = 0$
- $(D-1)-$ dim hyperplane, same form as the 2-class case: analogous properties
- The decision region for a multi-class linear discriminant must be convex & singly connected
- Enforced by the formulation! How?

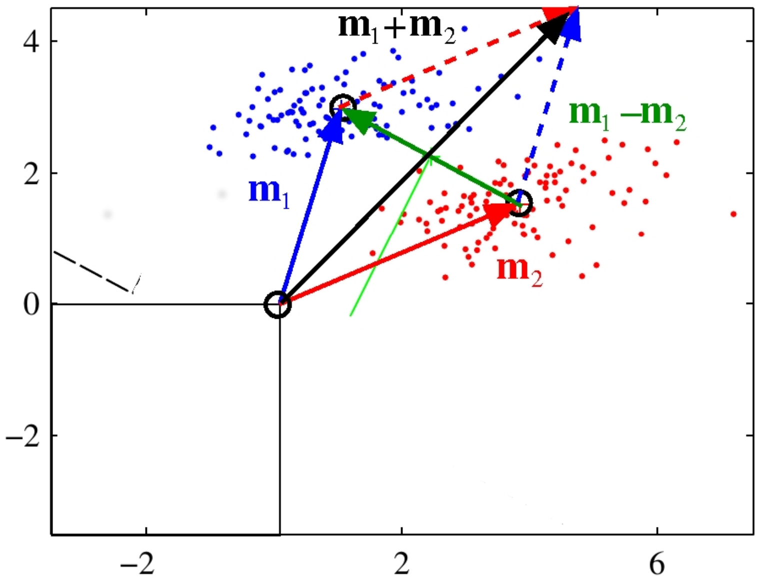


- $\hat{\mathbf{x}}$ lies on the line b/w $\mathbf{x}_A$ & $\mathbf{x}_B$
- $\hat{\mathbf{x}} = \lambda \mathbf{x}_A + (1-\lambda)\mathbf{x}_B$
- Discriminant Fn Convexity
- $y_j(\hat{\mathbf{x}}) = \mathbf{w}_j^T \hat{\mathbf{x}} + w_{j0} = \mathbf{w}_j^T(\lambda \mathbf{x}_A + (1-\lambda)\mathbf{x}_B) + w_{j0} =$
- $\lambda \mathbf{w}_j^T \mathbf{x}_A + (1-\lambda)\mathbf{w}_j^T \mathbf{x}_B + w_{j0} = \lambda[\mathbf{w}_j^T \mathbf{x}_A + w_{j0}] + (1-\lambda)[\mathbf{w}_j^T \mathbf{x}_B + w_{j0}] + w_{j0} - \lambda w_{j0} - (1-\lambda) w_{j0}$
- $\Longrightarrow$ $\boxed{y_j(\hat{\mathbf{x}}) = \lambda y_j(\mathbf{x}_A) + (1-\lambda) y_j(\mathbf{x}_B)}$

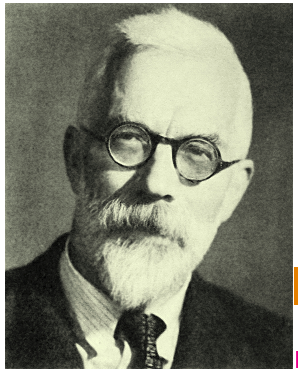
Some Physical Significance



- Points $\mathbf{x}$ in 2-D space with means $\mathbf{m}_1$ & $\mathbf{m}_2$ (2 classes)
- A line: $w_2 x_2 + w_1 x_1 + w_0 = 0$: $\mathbf{w}^T \mathbf{x} + w_0 = 0$
- $\mathbf{w} \perp \mathbf{x}$, perp dist from [0, 0] = $\frac{w_0}{||\mathbf{w}^T \mathbf{w}||}$
- All 2-D spaces can be superimposed (Euclidean, here): $[x_2\ x_1]^T$ or $[w_2\ w_1]^T$
- $\mathbf{w}^T \mathbf{x}$: all points $\mathbf{x}$ are projected onto line $\mathbf{w}$
- Line-Point Duality. Line $\mathbf{w}$: by intercepts w_2 & w_1
- Means (2-D points) $\mathbf{m}_1$ & $\mathbf{m}_2$ are projected to 1-D projections m_1 & m_2. Each point $\mathbf{x}$ to x



- 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 $\mathbf{m}_1 + \mathbf{m}_2$



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

Home Page

Title Page

Contents

Page 27 of 30

Go Back

Full Screen

Close

Quit

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]