3.9.2 Fisher discriminant

- First pattern recognition algorithm ever used (1936)
- This is equivalent to Bayes method (with Gaussian PDFs and equal covariances).
  - This simple definition is not well known—people used it for years without realizing this characteristic

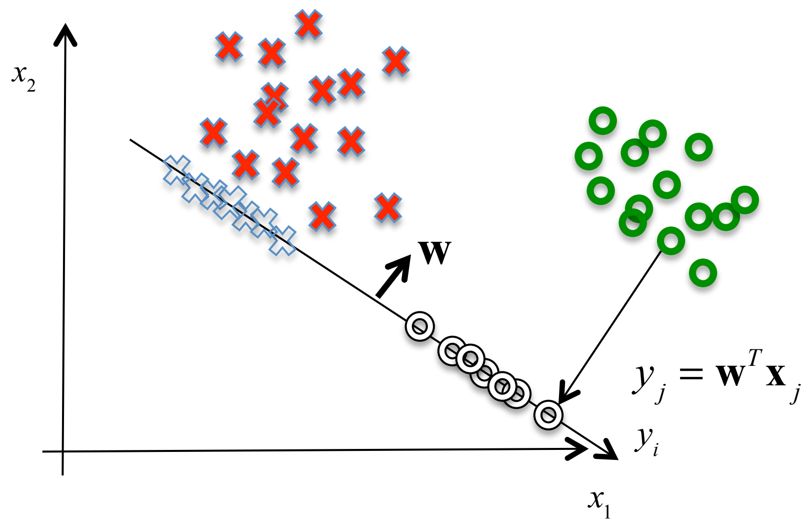
- Uses 1st and 2nd order statistics of PDF:

Mean: $$\mathbf{m}_i = \frac{1}{N_i} \sum_{\mathbf{x}_j \in w_i} \mathbf{x}_j$$

"Scatter": $$S = \frac{1}{2}(S_1 + S_2)$$

$$S_i = \sum_{\mathbf{x}_j \in W_i} (\mathbf{x}_j - \mathbf{m}_i)(\mathbf{x}_j - \mathbf{m}_i)^T = N_i \operatorname{cov}(\mathbf{x}_j \mid w_i)$$

- For a successful separation of classes, we want to LINEARLY project data onto scalar values y that can separate two classes with biggest gap between classes as possible.



- We can find the mean and scatter values of the clusters projections:

$$\tilde{\mathbf{m}}_i = \mathbf{w}^T \mathbf{m}_i$$

$$\tilde{S}_i = \mathbf{w}^T S_i \mathbf{w}$$

- Now lets define objective function for good separation:
- $J(\mathbf{w})$= (Between-class-scatter = distance between means) / (within-class-scatter) that is :

$$J(\mathbf{w}) = \frac{\left(\tilde{\mathbf{m}}_1 - \tilde{\mathbf{m}}_2\right)^2}{\tilde{S}_1 + \tilde{S}_2} = \frac{\mathbf{w}^T\left(\mathbf{m}_1 - \mathbf{m}_2\right)\left(\mathbf{m}_1 - \mathbf{m}_2\right)^T \mathbf{w}}{\mathbf{w}^T\left(S_1 + S_2\right)\mathbf{w}}$$

$$\mathbf{w}^* = \arg\min_{\mathbf{w}} J(\mathbf{w})$$

- Minimization of this problems can be transformed to Eigen value problem shown by using Schwarz inequality to have solution in form of :

$$\mathbf{w}^* = S^{-1}\left(\mathbf{m}_1 - \mathbf{m}_2\right)$$

which is the same as 2 class case in 3.8.1 for normal PDF with equal covariance for each class.