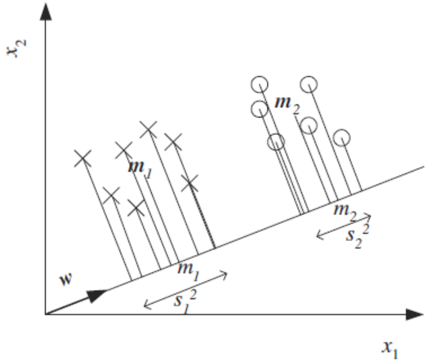
Compressing Data (Special Case) I

(Adapted from Alpaydin, *Introduction to Machine Learning*)

- Consider the special case when the data are known beforehand to be from some groups (or classes). We will only look at the case were data are from 2 classes (dependent variable values"1" and "0"). In general, each data point is a

Compressing Data (Special Case) II

point in d-
dimensional space. A two-dimensional example is shown below:



Compressing Data (Special Case) III

- In Linear Discriminant Analysis (LDA): not to be confused with Latent Dirichlet Allocation), we want to find a direction $\mathbf{w}$ s.t. when the data are projected onto $\mathbf{w}$ they are well separated
- LDA thus reduces the d-dimensional data to 1 dimension that allows the best separation into the classes
- The projection of a vector $\mathbf{x}$ onto a vector $\mathbf{w}$ is

$$\mathbf{z} = \frac{1}{|\mathbf{w}|}\mathbf{w}^T\mathbf{x}$$

- In the figure m_1 is the projection of $\mathbf{m}_1$ and m_2 is the projection of $\mathbf{m}_2$ onto $\mathbf{w}$. We can take $\mathbf{w}$ to be a unit-vector, since we are only interested in the magnitude of the projections in direction (angle) of $\mathbf{w}$ That is:

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

Compressing Data (Special Case) IV

- Assuming Class 1 is represented by $y = 1$ and Class 2 is represented by $y = 0$, then, given data with N instances:

$$m_1 = \frac{\sum_{i=1}^{N} \mathbf{w}^T \mathbf{x}_i y_i}{\sum_{i=1}^{N} y_i}$$

and:

$$m_2 = \frac{\sum_{i=1}^{N} \mathbf{w}^T \mathbf{x}_i (1 - y_i)}{\sum_{i=1}^{N} (1 - y_i)}$$

- The scatter of the projected instances about their (projected) means is:

$$s_1^2 = \sum_{i=1}^{N} (\mathbf{w}^T \mathbf{x}_i - m_1)^2 y_i$$

and:

$$s_2^2 = \sum_{i=1}^{N} (\mathbf{w}^T \mathbf{x}_i - m_1)^2 (1 - y_i)$$

Compressing Data (Special Case) V

- We want to find a **w** s.t. m_1 and m_2 are as far apart as possible, and s_1^2 and s_2^2 are as small as possible. That is, we want to maximize:

$$f(\mathbf{w}) = \frac{(m_1 - m_2)^2}{s_1^2 + s_2^2}$$

- The numerator is:

$$\begin{aligned} (m_1 - m_2)^2 &= (\mathbf{w}^T\mathbf{m}_1 - \mathbf{w}^T\mathbf{m}_2)^2 \\ &= \mathbf{w}^T(\mathbf{m}_1 - \mathbf{m}_2)(\mathbf{m}_1 - \mathbf{m}_2)^T\mathbf{w} \\ &= \mathbf{w}^T\mathrm{S}_{1,2}\mathbf{w} \end{aligned}$$

Compressing Data (Special Case) VI

and the denominator is:

$$\begin{aligned} s_1^2 &= \sum_{i=1}^{N} (\mathbf{w}^T \mathbf{x}_i - m_1)^2 \\ &= \sum_i \mathbf{w}^T (\mathbf{x}_i - \mathbf{m}_1)(\mathbf{x}_i - \mathbf{m}_1)^T \mathbf{w} \\ &= \mathbf{w}^T \mathrm{S}_1 \mathbf{w} \end{aligned}$$

where $\mathrm{S}_1 = \sum_i (\mathbf{x}_i - \mathbf{m}_1)(\mathbf{x}_i - \mathbf{m}_1)^T$.
Similarly:

$$s_2^2 = \mathbf{w}^T \mathbf{S}_2 \mathbf{w}$$

and the denominator is:

$$s_1^2 + s_2^2 = \mathbf{w}^T (\mathrm{S}_1 + \mathrm{S}_2) \mathbf{w}$$

Compressing Data (Special Case) VII

- Let $\mathrm{S} = \mathrm{S}_1 + \mathrm{S}_2$. Recall we want to minimise:

$$f(\mathbf{w}) \;=\; \frac{\mathbf{w}^T \mathrm{S}_{1,2} \mathbf{w}}{\mathbf{w} \mathrm{S} \mathrm{w}}$$

- It can be shown that if the partial derivation of f w.r.t. $\mathbf{w}$ is 0 then:

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