Tik-61.231 Principles of Pattern Recognition

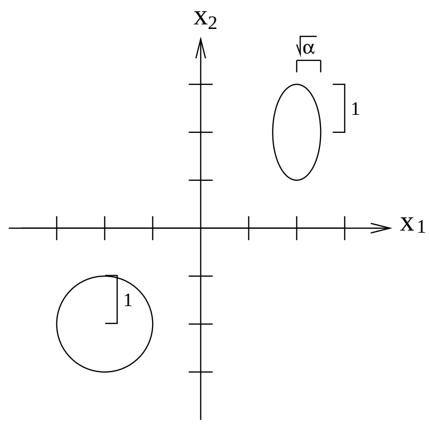
Answers to exercise 4: 16.10.2000

1. A Gaussian distribution has two parameters: a mean vector $\overline{\mu}$ and covariance matrix Σ.

   $\hat{\overline{\mu_i}} = \overline{m_i}$ and $\hat{\Sigma} = \frac{1}{n_i - 1}\sum_{k=1}^{n_i}(\overline{x_k} - \hat{\overline{\mu_i}})(\overline{x_k} - \hat{\overline{\mu_i}})^T$

   a) Lets assume that there are approximately as many samples from both classes ($n_1 \approx n_2$). Now we can draw both distributions in the same picture and use the same scale for both of them.

   The density of class 1 is symmetric as the diagonal elements of S_1 are equal. As for class 2, the density is expanded in the direction of the "width" of the distribution on the x_1-axis depends on α.



   b) $\hat{\overline{w}} = S_W^{-1}(\overline{m_1} - \overline{m_2})$

$$S_W = S_1 + S_2 = \begin{bmatrix} 1+\alpha & 0 \\ 0 & 1 \end{bmatrix} \Rightarrow S_W^{-1} = \frac{1}{2(1+\alpha)}\begin{bmatrix} 2+\alpha & 0 \\ 0 & 1+\alpha \end{bmatrix} = \begin{bmatrix} \frac{1}{1+\alpha} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}$$

$$\hat{\overline{w}} = \begin{bmatrix} \frac{1}{1+\alpha} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}\begin{bmatrix} (-2-2) \\ (-2-2) \end{bmatrix} = -2\begin{bmatrix} \frac{2}{1+\alpha} \\ 1 \end{bmatrix}$$

   c) To determine the eigenvectors of $S_W^{-1}S_B$ we first calculate the between-class scatter matrix S_B from $S_B = (\overline{m_1} - \overline{m_2})(\overline{m_1} - \overline{m_2})^T$.

$$S_B = \begin{bmatrix} (-2-2) \\ (-2-2) \end{bmatrix}\begin{bmatrix} (-2-2) & (-2-2) \end{bmatrix} = \begin{bmatrix} 16 & 16 \\ 16 & 16 \end{bmatrix}$$

   Now

$$S_W^{-1}S_B = \begin{bmatrix} \frac{1}{1+\alpha} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}\begin{bmatrix} 16 & 16 \\ 16 & 16 \end{bmatrix} = \begin{bmatrix} \frac{16}{1+\alpha} & \frac{16}{1+\alpha} \\ 8 & 8 \end{bmatrix}$$

   The eigenvectors for matrix A are defined as $A\overline{x} = \lambda\overline{x} \Rightarrow (A - \lambda I)\overline{x} = \overline{0}$. A nontrivial solution exists, if $det(A - \lambda I) = 0$

$$\begin{vmatrix} \frac{16}{1+\alpha} - \lambda & \frac{16}{1+\alpha} \\ 8 & 8-\lambda \end{vmatrix} = 0 \Leftrightarrow (\frac{16}{1+\alpha} - \lambda)(8-\lambda) - 8\frac{16}{1+\alpha} = 0 \Leftrightarrow \lambda^2 - (8 + \frac{16}{1+\alpha})\lambda = 0$$

Thus the eigenvalues are $\lambda_1 = 0$ and $\lambda_2 = 8 + \frac{16}{1+\alpha}$ (Note that $\alpha = \sigma_1^2 > 0$).
The eigenvector corresponding to the larger eigenvalue: $A\overline{e} = \lambda_2 \overline{e}$, where $\overline{e} = [e_1\ e_2]^T$. Thus

$$\begin{bmatrix} \frac{16}{1+\alpha} & \frac{16}{1+\alpha} \\ 8 & 8 \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} = (8 + \frac{16}{1+\alpha}) \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$$

Calculating from the lower row (the same result can be obtained from the upper row, too)

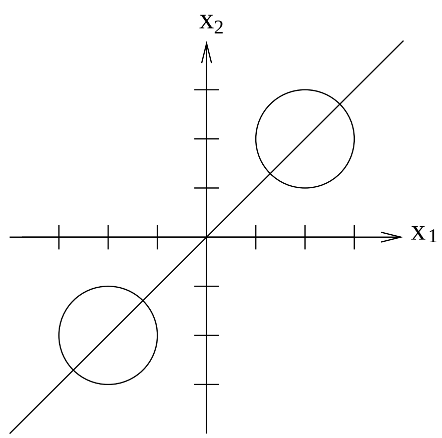
$$8e_1 + 8e_2 = (8 + \frac{16}{1+\alpha})e_2 \Leftrightarrow e_1 = \frac{2}{1+\alpha} e_2 \Rightarrow e = \begin{bmatrix} \frac{2}{1+\alpha} \\ 1 \end{bmatrix}$$

As can be seen, both methods produced a vector of the same orientation, as was to be expected.

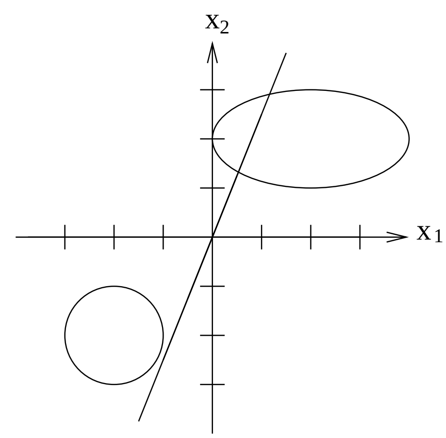
d) The Fisher linear discriminant is thus $\hat{\overline{w}} = \begin{bmatrix} \frac{2}{1+\alpha} \\ 1 \end{bmatrix}$, $\underline{m}_1 = (\ -2 \quad -2\ )^T$, $\underline{m}_2 = (\ 2 \quad 2\ )^T$,

$\mathbf{S}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $\mathbf{S}_2 = \begin{bmatrix} \alpha & 0 \\ 0 & 1 \end{bmatrix}$.

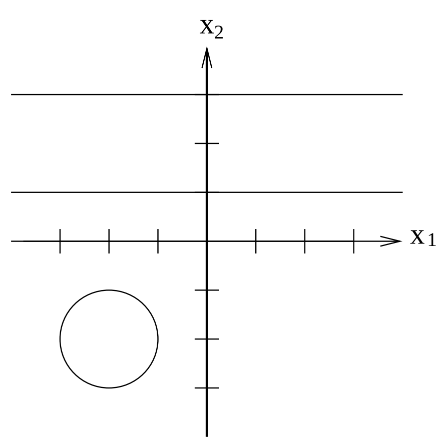
$\alpha = 1 \Rightarrow \hat{\overline{w}} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $S_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$:



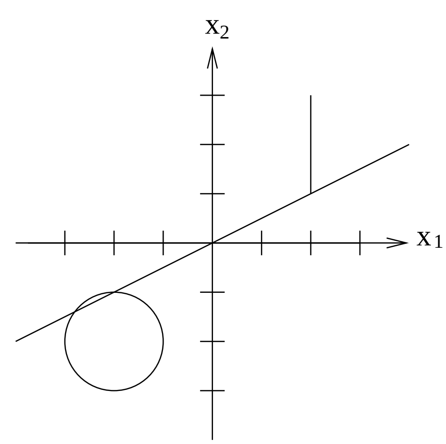
$\alpha = 4 \Rightarrow \hat{\overline{w}} = \begin{bmatrix} 2/5 \\ 1 \end{bmatrix}$ and $S_2 = \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}$:



$\alpha \to \infty \Rightarrow \hat{\overline{w}} \to \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ and $S_2 \to \begin{bmatrix} \infty & 0 \\ 0 & 1 \end{bmatrix}$:



$\alpha \to 0 \Rightarrow \hat{\overline{w}} \to \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ and $S_2 \to \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$:



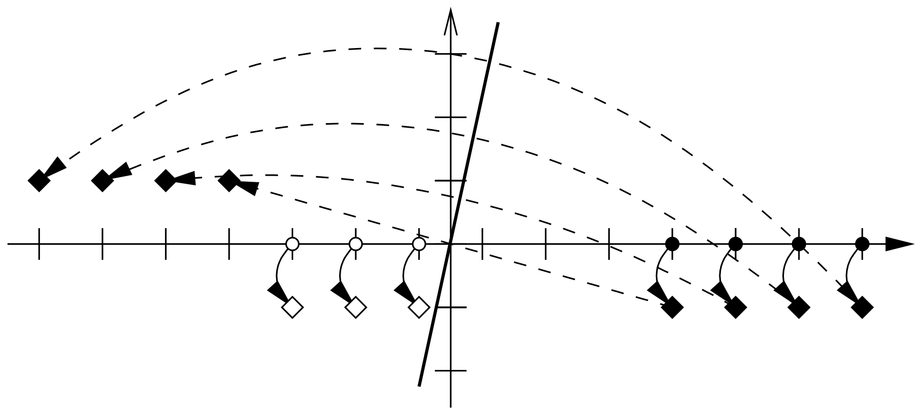
The discriminants seem quite valid, as it can be seen that by projecting the distributions onto the discriminant in all cases the distributions become well separable.

2. $\underline{\omega}^T \underline{x} - \omega_0 \begin{cases} > 0, \forall \underline{x} \in H_1 \\ < 0, \forall \underline{x} \in H_2 \end{cases}$.

a) $\underline{\omega}'^T \underline{x} - \omega_0 = \begin{bmatrix} \underline{\omega}' & \omega_0 \end{bmatrix} \begin{bmatrix} \underline{x} \\ -1 \end{bmatrix} = \underline{\omega}^T \underline{\hat{x}}$

If $\underline{\hat{x}} \in H_1 \Rightarrow \underline{\omega}^T \underline{\hat{x}} > 0$. If $\underline{\hat{x}} \in H_2 \Rightarrow \underline{\omega}^T \underline{\hat{x}} < 0 \Rightarrow \underline{\omega}^T(-1)\underline{\hat{x}} > 0$. So if the sign of $\underline{\hat{x}}$ is changed when $\underline{\hat{x}} \in H_2$, it holds that $\underline{\omega}^T \underline{\hat{x}} > 0 \; \forall \underline{\hat{x}}$

This is illustrated in the figure below, where the arrows represent the augmentation and dashed arrow the change in sign. After the augmentation and sign change all the samples are on the same side of the line.



3. $\underline{\omega}^{(n)^T} \underline{\hat{x}}_i < 0$ means that $\underline{\hat{x}}_i$ was not correctly classified in the nth iteration round.

In order to have $\underline{\hat{x}}_i$ correctly classified in the next iteration, $\underline{\omega}^{(n)}$ must be updated so that $\underline{\omega}^{(n+1)^T} \underline{\hat{x}}_i > 0$
The updating rule is $\underline{\omega}^{(n+1)} = \underline{\omega}^{(n)} + \alpha \underline{\hat{x}}_i$

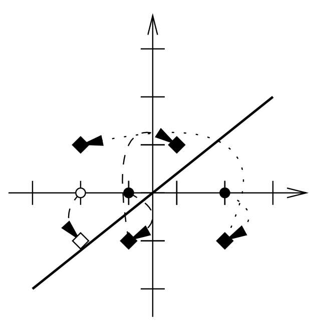
$$\Rightarrow \underline{\omega}^{(n+1)^T} \underline{\hat{x}}_i = (\underline{\omega}^{(n)} + \alpha \underline{\hat{x}}_i)^T \underline{\hat{x}}_i = \underline{\omega}^{(n)^T} \underline{\hat{x}}_i + \alpha_n ||\underline{\hat{x}}_i||^2 > 0 \Leftrightarrow \alpha_n > \frac{-\underline{\omega}^{(n)^T} \underline{\hat{x}}_i}{||\underline{\hat{x}}_i||^2}$$

4. Let's augment the vectors and choose to change the sign of the samples belonging to H_2. Thus the sample vectors are $\underline{x}_1 = \begin{bmatrix} -2 \\ -1 \end{bmatrix} \in H_1$ and $\underline{x}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \in H_2$, $\underline{x}_3 = \begin{bmatrix} -2 \\ 1 \end{bmatrix} \in H_2$. The initial weight vector is $\underline{\omega}^0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \in H_1$ and $\alpha = 0.5$.

The iteration will proceed as follows: (Note that the update is made even when $\underline{\omega}^T \underline{x} = 0$, even though this is not really indicated by the rule. I think that it is better to update them too often as this should only improve convergence, and as the original decision rule is based on < 0 or > 0 the situation $\underline{\omega}^T \underline{x} = 0$ is "indecisive". Although, in real-life situations the values would probably never be so precise as for 0 to be the result and this slight dilemma would never be seen.)

Sample	$\underline{\omega}^T \underline{x}$	New $\underline{\omega}^T$
1	$-2-1=-3<0$	$[(1-1)\ \ (1-0.5)]=[0\ \ 0.5]$
2	$0+0.5=0.5>0$	no change
3	$0+0.5=0.5>0$	no change
1	$0-0.5=-0.5<0$	$[(0-1)\ \ (0.5-0.5)]=[-1\ \ 0]$
2	$-1+0=-1<0$	$[(-1+0.5)\ \ (0+0.5)]=[-0.5\ \ 0.5]$
3	$1+0.5=1.5>0$	no change
1	$1-0.5=0.5>0$	no change
2	$-0.5+0.5=0$	$[(-0.5+0.5)\ \ (0.5+0.5)]=[0\ \ 1]$
3	$0+1=1>0$	no change
1	$0-1=-1<0$	$[(0-1)\ \ (1-0.5)]=[-1\ \ 0.5]$
2	$-1+0.5=-0.5<0$	$[(-1+0.5)\ \ (0.5+0.5)]=[-0.5\ \ 1]$
3	$1+1=2>0$	no change
1	$1-1=0$	$[(-0.5-1)\ \ (1-0.5)]=[-1.5\ \ 0.5]$
2	$-1.5+0.5=-1<0$	$[(-1.5+0.5)\ \ (0.5+0.5)]=[-1\ \ 1]$
3	$2+1=3>0$	no change
1	$2-1=1>0$	no change
2	$-1+1=0$	$[(-1+0.5)\ \ (1+0.5)]=[0.5\ \ 1.5]$
3	$-1+1.5=0.5>0$	no change
1	$1-1.5=-0.5<0$	$[(-0.5-1)\ \ (1.5-0.5)]=[-1.5\ \ 1]$
2	$-1.5+1=0.5<0$	$[(-1.5+0.5)\ \ (1.5+0.5)]=[-1\ \ 1.5]$
3	$2+1.5=2.5>0$	no change
1	$2-1.5=0.5>0$	no change
2	$-1+1.5=0.5>0$	no change

So the algorithm converged at a weight vector of $\underline{\omega}=[-1\ \ 1.5]^T=[\omega'\ \ \omega_0]$, which is illustrated in the figure below.



The correctness of the result can also be verified by calculating with the original values:

$$x_1=-2 \Rightarrow \omega' x-\omega_0=-1*-2-1.5=0.5>0$$
$$x_2=-1 \Rightarrow \omega' x-\omega_0=-1*-1-1.5=-0.5<0$$
$$x_3=2 \Rightarrow \omega' x-\omega_0=-1*2-1.5=-3.5<0$$