HW#2 Solutions, EE556

1) Recall: a linear machine chooses

$$\hat{c}(\underline{x}) = \arg\max_j g_j(\underline{x}), \qquad (*)$$

where $g_j(\underline{x}) = \underline{w}_j^T \underline{x} + w_{j0}$.

Now, suppose two points $\underline{x}_0$ and $\underline{x}_1$ both get assigned, via $(*)$, to the same class, e.g. class i. A linear machine produces convex decision regions if $\lambda \underline{x}_0 + (1-\lambda)\underline{x}_1$ <u>also</u> gets assigned to class i, for all $0 \le \lambda \le 1$. (Note: the class, i, was arbitrarily chosen...).

Now, from the assumption, we have:

$$\max_j g_j(\underline{x}_0) = g_i(\underline{x}_0) = \underline{w}_i^T \underline{x}_0 + w_{i0}$$

$$\max_j g_j(\underline{x}_1) = g_i(\underline{x}_1) = \underline{w}_i^T \underline{x}_1 + w_{i0}$$

Next, consider

$$g_j(\lambda \underline{x}_0 + (1-\lambda)\underline{x}_1) = \underline{w}_j^T(\lambda \underline{x}_0 + (1-\lambda)\underline{x}_1) + w_{j0}$$
$$= \lambda(\underline{w}_j^T \underline{x}_0 + w_{j0}) + (1-\lambda)(\underline{w}_j^T \underline{x}_1 + w_{j0}).$$

Clearly,

$$\max_j g_j(\lambda \underline{x}_0 + (1-\lambda)\underline{x}_1) \le \lambda \max_j g_j(\underline{x}_0) + (1-\lambda)\max_j g_j(\underline{x}_1) \qquad (\Delta)$$

But

$$g_i(\lambda \underline{x}_0 + (1-\lambda)\underline{x}_1) = \lambda(\underline{w}_i^T \underline{x}_0 + w_{i0}) + (1-\lambda)(\underline{w}_i^T \underline{x}_1 + w_{i0})$$
$$= \lambda \max_j g_j(\underline{x}_0) + (1-\lambda)\max_j g_j(\underline{x}_1)$$

$\Rightarrow$ $\lambda \underline{x}_0 + (1-\lambda)\underline{x}_1$ is also assigned to class i by the linear machine.

$\therefore$, linear machines produce convex decision regions.

2) i) Let's solve $\min_{\underline{x}_q} \|\underline{x} - \underline{x}_q\|^2$ s.t. $g(\underline{x}_q) = 0$,

where $g(\underline{x}) = \underline{w}^T\underline{x} + w_0$.

We'll use the <u>method of Lagrange multipliers</u> from calculus:

We first form the Lagrangian cost function:

$$L = \|\underline{x} - \underline{x}_q\|^2 + \lambda(\underline{w}^T\underline{x}_q + w_0 - 0).$$

Then,

$$(1) \quad \nabla_{\underline{x}_q} L = 2(\underline{x}_q - \underline{x}) + \lambda\underline{w} = \underline{0}$$

and

$$(2) \quad \frac{\partial L}{\partial \lambda} = \underline{w}^T\underline{x}_q + w_0 = 0$$

} @ an extremum of the Lagrangian, L.

(1) $\Rightarrow$ $2\underline{x}_q = 2\underline{x} - \lambda\underline{w}$. Plug this into (2) $\Rightarrow$

$$\underline{w}^T\left(\frac{2\underline{x} - \lambda\underline{w}}{2}\right) + w_0 = 0 \Rightarrow$$

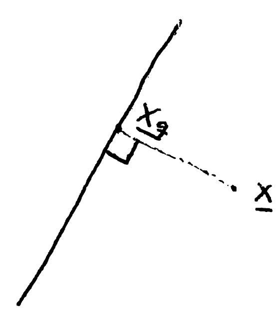
$$\underline{w}^T\underline{x} - \frac{\lambda}{2}\|\underline{w}\|^2 + w_0 = 0 \quad \text{or}$$

$$g(\underline{x}) - \frac{\lambda}{2}\|\underline{w}\|^2 = 0 \Rightarrow \lambda = \frac{2g(\underline{x})}{\|\underline{w}\|^2}$$

or, $$\underline{x}_q = \underline{x} - \frac{g(\underline{x})\underline{w}}{\|\underline{w}\|^2} \Rightarrow$$

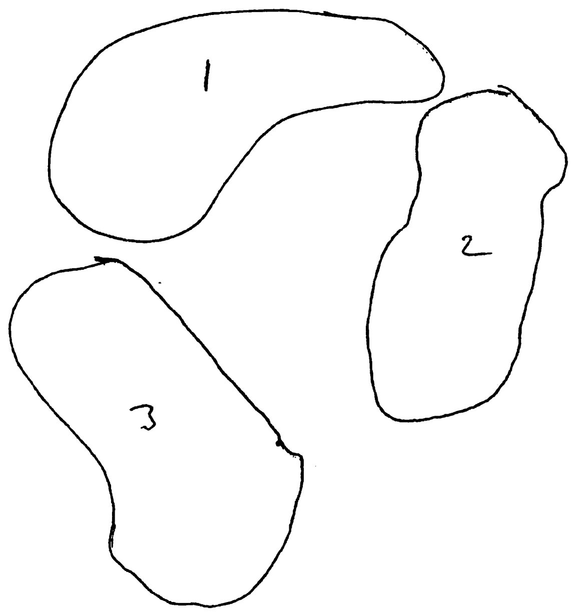
$$\|\underline{x}_q - \underline{x}\|^2 = \frac{|g(\underline{x})|^2\underline{w}^T\underline{w}}{\|\underline{w}\|^4}$$

$$\sqrt{} \quad = \frac{|g(\underline{x})|}{\|\underline{w}\|}.$$



ii) We already found the point on the decision boundary in part i): $\underline{x}_q = \underline{x} - \dfrac{g(\underline{x})\underline{w}}{\|\underline{w}\|^2}$

3) Pretty easy to illustrate this graphically. Consider 3 classes in the plane:



These classes <u>are</u> pairwise linearly separable, but you cannot, e.g., linearly separate class 1 from classes 2 and 3.

4) Let $b = \underline{x}^T \underline{1}$, where $\underline{1} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$ (all ones vector).

Then, $W_1 = \left\{ \underline{x} : b \text{ is odd} \right.$

$W_2 = \left\{ \underline{x} : b \text{ is even} \right.$

a) Let's show these classes are <u>not</u> lin. sep. by showing a contradiction if we assume it to be true.

Spse. $\exists\ (\underline{w}, w_0)$ s.t.

$$\underline{w}^T \underline{x} + w_0 \geq 0 \quad \text{for } \underline{x} \in W_1$$

and

$$\underline{w}^T \underline{x} + w_0 < 0 \quad \text{for } \underline{x} \in W_2$$

$\underline{x} = \underline{0} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \Rightarrow b = 0$ + belongs to W_1; $\therefore$,

$\underline{x} = \underline{0}$ must satisfy $w_0 \geq 0$ (1)

$\underline{x} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$, i.e. a vector with one '1', $\Rightarrow b = 1$ + belongs to W_2.

$\therefore$, it must satisfy $\underbrace{w_i + w_0 < 0}_{(2)}$, i the position of the '1'.

$\underline{x} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$, i.e. a vector with two '1's $\Rightarrow b = 2$ + belongs to W_1.

Suppose the ones are in positions i and j, i.e.

(3) $\quad w_i + w_j + w_0 \geq 0, \ i \neq j.$

Consider eqn. (2) for positions i and j: $\quad w_i + w_0 < 0$

$w_j + w_0 < 0$

Also, rewrite (1) as $-w_0 \leq 0$

Adding these 3 eqns together, we get.

$$\underbrace{(w_i + w_0)}_{<0} + \underbrace{(w_j + w_0)}_{<0} + \underbrace{(-w_0)}_{\leq 0} < 0 \ !$$

But this <u>contradicts</u> (3).

b) For simplicity, let's assume d is even.

First, we observe (recall) that $\underline{w}' = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$ can be used to count the number of ones in $\underline{x}$, via $K_0 = \underline{w}'^T \underline{x}$.

This immediately suggests the following as a possible solution strategy:

1) Choose $\ell^* = \underset{\ell \in \{0,1,\dots d\}}{\arg\min} (\underline{w}'^T \underline{x} - \ell)^2$

2) Take the parity of ℓ^* as the decision result.

This method certainly works, but unfortunately is not based on linear discriminants, but rather quadratic discriminants (expand the square above to see this clearly).

However, something related that is based on linear discriminants will in fact work.

First, just consider the even integers 0, 2, 4, ... d.

$$|\underline{w}'^T \underline{x} - 2m|^2 = |K_0 - 2m|^2 = K_0^2 - 4mK_0 + 4m^2$$

$$\text{Clearly, } \underset{m \in \{0,1,\dots d/2\}}{\arg\min} |K_0 - 2m|^2 = \underset{m \in \{0,1,\dots d/2\}}{\arg\min} -4mK_0 + 4m^2$$

$$= \underset{m \in \{0,1,\dots d/2\}}{\arg\max} 4mK_0 - 4m^2$$

$$= \underset{m \in \{0,1,\dots d/2\}}{\arg\max} 4m\underline{w}'^T \underline{x} - 4m^2 \triangleq m_e^* \qquad (\square)$$

(obvious from above, but you can also take derivatives w.r.t. m, set to zero, & solve for m_e^*).

It's easy to verify in several ways that $m_e^* = \frac{K_0}{2}$, i.e. the even integer result is $2m_e^*$.

Further recognize that $(\square)$ specifies a set of $d/2+1$ linear discriminant functions, and a way of selecting the nearest even integer to K_0.