

1. In discussion we derived an expression for the signed distance  $d$  between an arbitrary point  $\underline{x}$  (or  $\underline{p}$ ) and a hyperplane  $H$  given by  $g(\underline{x}) = w_0 + \underline{w}^T \underline{x} = 0$ , all in *nonaugmented* feature space. This question explores this topic further.

- (a) Prove that the weight vector  $\underline{w}$  is normal to  $H$ .

**Hint:** For any two points  $\underline{x}_1$  and  $\underline{x}_2$  on  $H$ , what is  $g(\underline{x}_1) - g(\underline{x}_2)$ ? How can you interpret the vector  $(\underline{x}_1 - \underline{x}_2)$ ?

- (b) Show that the vector  $\underline{w}$  points to the positive side of  $H$ . (Positive side of  $H$  means the  $d > 0$  side.)

**Hint:** What sign does the distance  $d$  from  $H$  to  $\underline{x} = (\underline{x}_1 + a \underline{w})$  have, in which  $\underline{x}_1$  is a point on  $H$ ?

- (c) Derive, or state and justify, an expression for the signed distance  $r$  between an arbitrary point  $\underline{x}^{(+)}$  and a hyperplane  $g(\underline{x}^{(+)}) = \underline{w}^{(+T)} \underline{x}^{(+)} = 0$  in *augmented* feature space. Set up the sign of your distance so that  $\underline{w}$  points to the positive-distance side of  $H$ .
- (d) In *weight* space, using augmented quantities, derive an expression for the signed distance between an arbitrary point  $\underline{w}^{(+)}$  and a hyperplane  $g(\underline{x}^{(+)}) = \underline{w}^{(+T)} \underline{x}^{(+)} = 0$ , in which the vector  $\underline{x}^{(+)}$  defines the positive side of the hyperplane.

2. For a 2-class learning problem with one feature, you are given three training data points (in augmented space):

$$\underline{x}_1^{(1)} = (1, -3); \underline{x}_2^{(1)} = (1, -5); \underline{x}_3^{(2)} = (1, 1); \underline{x}_4^{(2)} = (1, -1)$$

- (a) Plot the data points in 2D feature space. Draw a linear decision boundary  $H$  that correctly classifies them, showing which side is positive.
- (b) Plot the reflected data points in 2D feature space. Draw the same decision boundary; does it still classify them correctly?
- (c) Plot the reflected data points, as lines in 2D weight space, showing the positive side of each. Show the solution region.
- (d) Also, plot the weight vector  $\underline{w}$  of  $H$  from part (a) as a point in weight space. Is  $\underline{w}$  in the solution region?

3. (a) Let  $p(\underline{x})$  be a scalar function of a  $D$ -dimensional vector  $\underline{x}$ , and  $f(p)$  be a scalar function of  $p$ . Prove that:

$$\nabla_{\underline{x}} f[p(\underline{x})] = \left[ \frac{d}{dp} f(p) \right] \nabla_{\underline{x}} p(\underline{x})$$

*i.e.*, prove that the chain rule applies in this way. [**Hint:** you can show it for the  $i^{\text{th}}$  component of the gradient vector, for any  $i$ . It can be done in a couple lines.]

- (b) Use relation (18) of DHS A.2.4 to find  $\nabla_{\underline{x}} (\underline{x}^T \underline{x})$ .
  - (c) Prove your result of  $\nabla_{\underline{x}} (\underline{x}^T \underline{x})$  in part (b) by, instead, writing out the components.
  - (d) Use (a) and (b) to find  $\nabla_{\underline{x}} \left[ (\underline{x}^T \underline{x})^3 \right]$  in terms of  $\underline{x}$ .
4. (a) Use relations above to find  $\nabla_{\underline{w}} \|\underline{w}\|_2$ . Express your answer in terms of  $\|\underline{w}\|_2$  where possible. **Hint:** let  $p = \underline{w}^T \underline{w}$ ; what is  $f$ ?
- (b) Find:  $\nabla_{\underline{w}} \|\underline{M}\underline{w} - \underline{b}\|_2$ . Express your result in simplest form. **Hint:** first choose  $p$  (remember it must be a scalar).
5. **[Extra credit]** For  $C > 2$ , show that total linear separability implies linear separability, and show that linear separability doesn't necessarily imply total linear separability. For the latter, a counterexample will suffice.