

58147 Machine Learning (Spring 2005)

Exercise 1 (2 February)

1. Represent the Boolean function

$$f(\mathbf{x}) = x_1(x_2 + x_3(x_4 + x_5x_6))$$

as (a) 1-decision list (1-DL), (b) DNF formula and (c) CNF formula.

2. Consider the monotone conjunction $g(\mathbf{x}) = x_1x_2 \dots x_k$ over $\{-1, 1\}^n$, where $1 \leq k \leq n$.

- (a) Represent g as a linear classifier so that the normalized margin

$$\min \left\{ \frac{y(\mathbf{w} \cdot \mathbf{x} - b)}{\|\mathbf{w}\|_2} \mid \mathbf{x} \in \{-1, 1\}^n, y = g(\mathbf{x}) \right\}$$

is as large as possible.

- (b) Represent g as a linear classifier so that $\|\mathbf{w}\|_2 = 1$ and the unnormalised margin

$$\min \{ y(\mathbf{w} \cdot \mathbf{x} - b) \mid \mathbf{x} \in \{-1, 1\}^n, y = g(\mathbf{x}) \}$$

is as large as possible.

- (c) Represent g as a linear classifier so that the unnormalized margin is exactly 1 and $\|\mathbf{w}\|_2$ is as small as possible.

- (d) Represent g in $n + 1$ dimensions so that $b = 0$ (Dietterich's "unthresholded hypothesis"). Compare the margin to part (a).

Note: the normalized margin is the minimum distance between the instances $\mathbf{x} \in \{-1, 1\}^n$ and the separating hyperplane $\mathbf{w} \cdot \mathbf{x} = b$. Solution to (a) is unique up to a constant factor, and (b) and (c) are obtained by choosing the constant factor appropriately. You don't need to prove that your classifiers have the stated optimality properties, as long as you understand the intuition.

3. Define the usual *exclusive or* (XOR) operation for bits $x, y \in \{-1, 1\}$ by $x \oplus y = 1$ if $x \neq y$, and $x \oplus y = -1$ if $x = y$.

- (a) Represent the 3 bit parity function $x_1 \oplus x_2 \oplus x_3$ as (a) decision tree and (b) DNF formula.

- (b) Show that the concept $x_1 \oplus x_2$ cannot be represented as a linear classifier.

4. Represent the Boolean function from Problem 1 as a linear classifier. What do you get as the normalised margin?

Hint: Think about the decision list representation. Choose weights such that $w_i = 2w_{i+1}$.

5. A function $f: X \rightarrow Y$ is *consistent* with a sample of examples $(x_t, y_t) \in X \times Y$, if $y_t = f(x_t)$ for all t .

Give an algorithm that takes as input a sample $((x_t, y_t))_{t=1}^m$, runs in time polynomial in the size of input, and outputs a conjunction that is consistent with the sample, assuming such a consistent conjunction exists.

Hint: consider the longest possible conjunction that is consistent with all the positive examples.