1: Warm Up: Feature Expansion

The concept class C consisting of functions f_r is defined by a radius r as follows:

$$f_r(x_1, x_2) = \begin{cases} +1 & 4x_1^4 + 16x_2^4 \leq r; \\ -1 & \text{otherwise} \end{cases}$$

This hypothesis class is *not* linearly separable in $\mathbb{R}^2$. To make positive and negative examples linearly separable, the examples must be mapped to a new space using a function $\phi(x_1, x_2)$ defined as :

$$\phi(x_1, x_2) = \begin{bmatrix} x_1^4 \\ x_2^4 \end{bmatrix}$$

To prove that the positive and negative points are linearly separated in this new space, we can produce a hyperplane that splits them. That is, a weight vector $\mathbf{w}$ and a bias b are found such that $\mathbf{w}^T\phi(x_1, x_2) \geq b$ if, and only if, $f_r(x_1, x_2) = +1$.

$$\mathbf{w} = \begin{bmatrix} -4 \\ -16 \end{bmatrix} \; and \; b = -r$$

2: Mistake Bound Model of Learning

(1) Each function f_r in a concept class **C** is defined by a radius r, where $1 \leq r \leq 80$. This gives the functions $f_1, f_2, ..., f_{79}, f_{80}$ in **C**. For a concept class of size 80.

$\boxed{|C| = 80}$

(2) Given an input point (x_1^t, x_2^t) along with it's label y^t, we can use the following expression to check whether the current hypothesis f_r has made a mistake.

$$\boxed{sgn((x_1^t)^2 + (x_2^t)^2 - r^2 - 1) = sgn(y^t)}$$

If both sides of the expression have the same sign, we know we have made a mistake. The intuition is $x_1^2 + x_2^2 - r^2$ will be negative in the case that r^2 is greater than $x_1^2 + x_2^2$ (an incorrect label of -1 is also negative). But $x_1^2 + x_2^2 - r^2$ will be positive when r^2 is less than $x_1^2 + x_2^2$ (an incorrect label of +1 is also positive).

There is an edge case where $x_1^2 + x_2^2 = r^2$. The incorrect labeling is -1, but $sgn(x_1^2 + x_2^2 - r^2) \neq sign(-1)$ in this case. To account for this, one is subtracted from the left side of the equation.

(3) When there is an error, the radius r must be updated. If there is a mistake when $y^t = +1$ then r will be increased by one. Otherwise, if $y^t = -1$ then r will be decreased by one. The radius r is bounded by $1 \leq r \leq 80$ and the modifications to r must obey this.

$$\boxed{y^t = +1 : increase \; r \; by \; one.}$$

$$\boxed{y^t = -1 : decrease \; r \; by \; one.}$$

(4)

(5a)

(5b)

(5c)

3.1: The Perceptron Algorithm and its Variants

(1) Running the simple Perceptron algorithm on the data from $table2$, with a learning rate $r = 0.5$, produces the following weight vector.

$$\mathbf{w} = \begin{bmatrix} 0 \\ 0.5 \\ 0 \\ -0.5 \\ 1 \end{bmatrix}$$

With one pass of the Perceptron algorithm, four mistakes were made on the `table2` dataset.

$\boxed{\text{Four mistakes made.}}$

(2) Before training a binary classifier with the Perceptron algorithm, I ran some tests to determine the best learning rate r. For each learning rate, the examples from the file `a5a.train` were ran through the Perceptron algorithm once.

Learning Rate	Updates Made
1	92.86%

Table 1: Number of updates made during training for various learning rates.

(3)

(4)