Introduction to Machine Learning, Fall 2015 - Exercise session V

Rodion "rodde" Efremov, student ID 013593012

November 24, 2015

Problem 2 (3 points)

Exclusive or is a classical example used to demonstrate the limitations of linear classifiers. In the most basic setting, we have two binary inputs and a binary output. Using +1 and -1 as the binary values, the exclusive or function XOR is given by the following table:

x_1	x_2	XOR(x_1, x_2)
-1	-1	-1
-1	+1	+1
+1	-1	+1
+1	+1	-1

(a) Give a detailed proof for the fact that the exclusive or function cannot be represented as a linear classifier. That is, prove that there is no coefficient vector (w_0, w_1, w_2) such that

$$\text{XOR}(x_1, x_2) = \text{sign}(w_0 + w_1 x_1 + w_2 + x_2)$$

whould hold for all possible inputs (x_1, x_2).

(b) Show that if we include an extra input variable $x_1 x_2$, we can represent XOR as a linear classifier in terms of the modified inputs. That is, find a coefficient vector (w_0, w_1, w_2, w_3) such that

$$\text{XOR}(x_1, x_2) = sign(w_0 + w_1 x_1 + w_2 x_2 + w_3 x_1 x_2)$$

holds for all (x_1, x_2).

Solution to (a)

Define three points in $\mathbb{R}^3$:

A $(-1, -1, w_0 - w_1 - w_2)$,

B $(-1, +1, w_0 - w_1 + w_2)$,

C $(+1, -1, w_0 + w_1 - w_2)$.

We must have

$$w_0 - w_1 - w_2 < 0,$$
$$w_0 - w_1 + w_2 > 0,$$
$$w_0 + w_1 - w_2 > 0.$$

Next, let us find a plane in $\mathbb{R}^3$ that passes through points A, B and C: $z = ax + by + c$. Now,

$$\begin{cases} -a - b + c & = w_0 - w_1 - w_2 \\ -a + b + c & = w_0 - w_1 + w_2 \\ a - b + c & = w_0 + w_1 - w_2, \end{cases}$$

which simplifies to

$$\begin{cases} -a - b + c & = w_0 - w_1 - w_2 \\ 2c & = 2w_0, \end{cases}$$

so we have that $c = w_0$. Next,

$$\begin{cases} -a - b & = -w_1 - w_2 \\ -a + b & = -w_1 + w_2. \end{cases}$$

Which leads to $-2a = -2w_1$ and $a = w_1$, and finally $b = w_2$. Now our plane is defined as

$$z = w_1 x + w_2 y + w_3.$$

The point D has coordinates $(+1, +1, w_0 + w_1 + w_2)$. Since

$$w_0 - w_1 + w_2 > 0$$
$$w_0 + w_1 - w_2 > 0,$$

summing the above inequalities we obtain $w_0 > 0$. Also, since $w_0 - w_1 - w_2 < 0$, we have that $w_0 < w_1 + w_2$. So $w_0 + w_1 + w_2 > 0$ and D is "misclassified" (we should have had $w_0 + w_1 + w_2 < 0$.).

If we chose instead of the point D, we would get the same contradiction. The problem is that all four points do not lie on a same single plane.