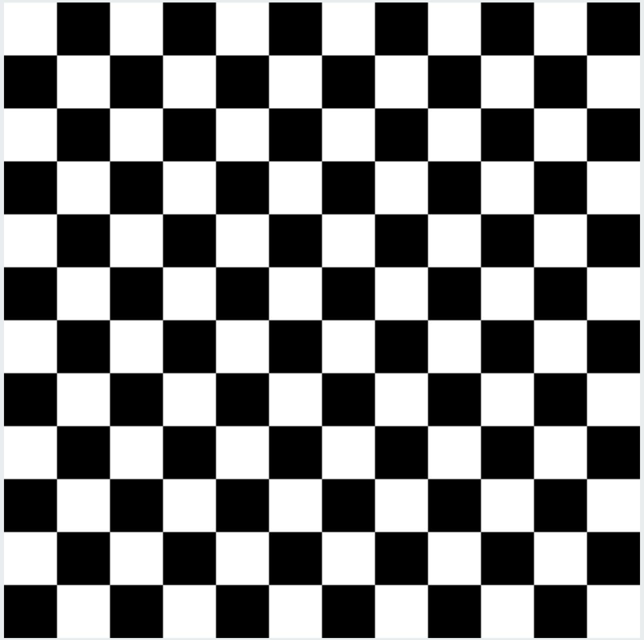
VC dimension of checkerboard patterns

Checkerboard patterns are hard to learn.

– anonymous colleague, 29.6.2018

Three questions arise:

1. What are checkerboard patterns?
2. What does it mean to learn a checkerboard pattern?
3. What does ‘hard’ mean?



What are checkerboard patterns?

In this talk we consider checkerboards of side length 2 with a total of $(2n)^2$ squares of side length $1/n$ each.

1-dimensional checkerboard patterns

Actually we will only look at the 1-dimensional version, since all ideas are the same but the math is simpler.

Notes:

- Each pattern corresponds to a natural number and vice versa.
- The intervals are left-closed, right-open

What does it mean to learn a checkerboard pattern?

Setting: We are given a sample $S = \{x_1,..., x_N\}$ with labels $\{y_1,..., y_N\}$ according to some unknown number n_0.

Goal: Determine the underlying checkerboard pattern.

Note: As each checkerboard pattern corresponds to a natural number, the question becomes: How can we reconstruct n_0 (or at least something similar) from S?

Example

The sample on the left is labeled according to some unknown number n_0.

For $n=2$ we get a model that fits the data. But $n=6$ also works!

If we choose the wrong n, how bad does our model generalize?

What does ‘hard’ mean?

Ignoring all sorts of technical details we have something like:

$$P(testErr \ < trainErr + \sqrt{D/N}) = 1\text{-}\varepsilon$$

where N is the sample size and D the VC dimension of the problem.

We see that a problem is harder if it has higher VC dimension.

VC dimension

The VC dimension is the largest d for which we can find a set $\{x_1,..., x_d\}$, such that any choice of labels $\{y_1,..., y_d\}$ can be realized by a checkerboard pattern, i.e. for any choice of labels there is a number n such that

$$\mathrm{floor}(n \cdot x_i) \,\%\, 2 = y_i.$$

Notes:

- For labels we use the interpretation: "black = 0", "white = 1"
- The VC dimension solely depends on the hypothesis space and "measures" the correlation of this space with random noise

The VC dimension is at least 3

For d=3 and $\{x_1, x_2, x_3\} = \{1, ½, ⅓\}$ we get the following table:

labels	(0, 0, 0)	(0, 0, 1)	(0, 1, 0)	(0, 1, 1)	(1, 0, 0)	(1, 0, 1)	(1, 1, 0)	(1, 1, 1)
n	0	4	2	10	1	5	7	3

Exercise: Check that the above does not work if we use the set $\{1, ⅓, ⅔\}$.

Another point of view

We can also think of the multiples

$$\mathrm{floor}(n \cdot x_i) \ \% \ 2$$

as points orbiting around the circle on the left.

Our intuition tells us that if the orbital speeds (i.e. the x_i) are different enough, the points will assume any position if we just wait long enough (i.e. if n is big enough).

What exactly is ‘different enough’?

Criterion 1 (via Kronecker Theorem):
The x_i are linearly independent over the rational numbers, e.g.

$$\{x_1, \ldots, x_d\} \subset \{1/\sqrt{p} \mid p \text{ prime}\}$$

Criterion 2 (via Chinese Remainder Theorem):
The x_i are non-integral reduced fractions a_i/b_i, with $\gcd(b_i, b_j) = 1$ and $\gcd(2, b_i) = 1$, e.g.

$$\{x_1, \ldots, x_d\} \subset \{1/p \mid p \text{ odd prime}\}$$

Conclusion

For each natural number d we can take the first d odd primes $p_1,...,p_d$ and the sequence $\{1/p_1, ..., 1/p_d\}$ to see that the VC dimension is at least d.

This shows that the VC dimension is infinite, which can with good reason be interpreted as ”Checkerboard patterns are hard to learn”.