Minsky and Papert's Proof

`*From 'Perceptrons: An Introduction to Computational Geometry' by Marvin Minsky and Seymour Papert,`

Suppose at step t, the perceptron with weight wt makes a mistake on data point (x, y), then it updates to wt+1 = wt + r(y - fwt(x))x.

If y = 0, the argument is symmetric, so we omit it.

WLOG, y = 1, then fwt(x) = 0, fw*(x) = 1, and wt+1 = wt + rx.

By assumption, we have separation with margins:

w* · x ≥ γ

Thus, w* · wt+1 - w* · wt = w* · (rx) ≥ rγ

Also

$||wt+1||2^{2 - ||wt||^{2}}2 = ||wt + rx||2^{2 - ||wt||^{2}}2 = 2r(wt \cdot x) + r^{2||x||^{2}}2$

and since the perceptron made a mistake, wt · x ≤ 0, and so

$||wt+1||2^{2 - ||wt||^{2}}2 \le ||x||2^{2 \le r}2R$^2

Since we started with w0 = 0, after making N mistakes,

$||w||2 \le \surd(Nr^{2R}2)$

but also

||w||2 ≥ w · w* ≥ Nrγ

Combining the two, we have N ≤ (R/γ)^2

Addendum to intial proof by Taylor Metz

`*Draft by Taylor Metz, 2025-01-31*`

The proof above assumes linear separability for the data point (x,y). However, for XOR patterns in the initial proof are considered not linear separable. Using the same weights and bias as the initial proof:

- Wx = |-1,0,+1|
- Wy = |-1,0,+1|
- bx = (0,1)
- by = (0,1)

The data point (x,y) sum to the percentage allocation XY where the summation is 2×100%.

Thus, {((xy)-1)×(-1/2)} resolves the logic to achieve separability.

- For x = -1, y = -1, the product xy is 1. Applying the transformation, (1 - 1) × (-1/2) results in 0.
- For x = -1, y = +1, the product xy is -1. Applying the transformation, (-1 - 1) × (-1/2) results in +1.
- For x = +1, y = -1, the product xy is -1. Applying the transformation, (-1 - 1) × (-1/2) results in +1.
- For x = +1, y = +1, the product xy is 1. Applying the transformation, (1 - 1) × (-1/2) results in 0.