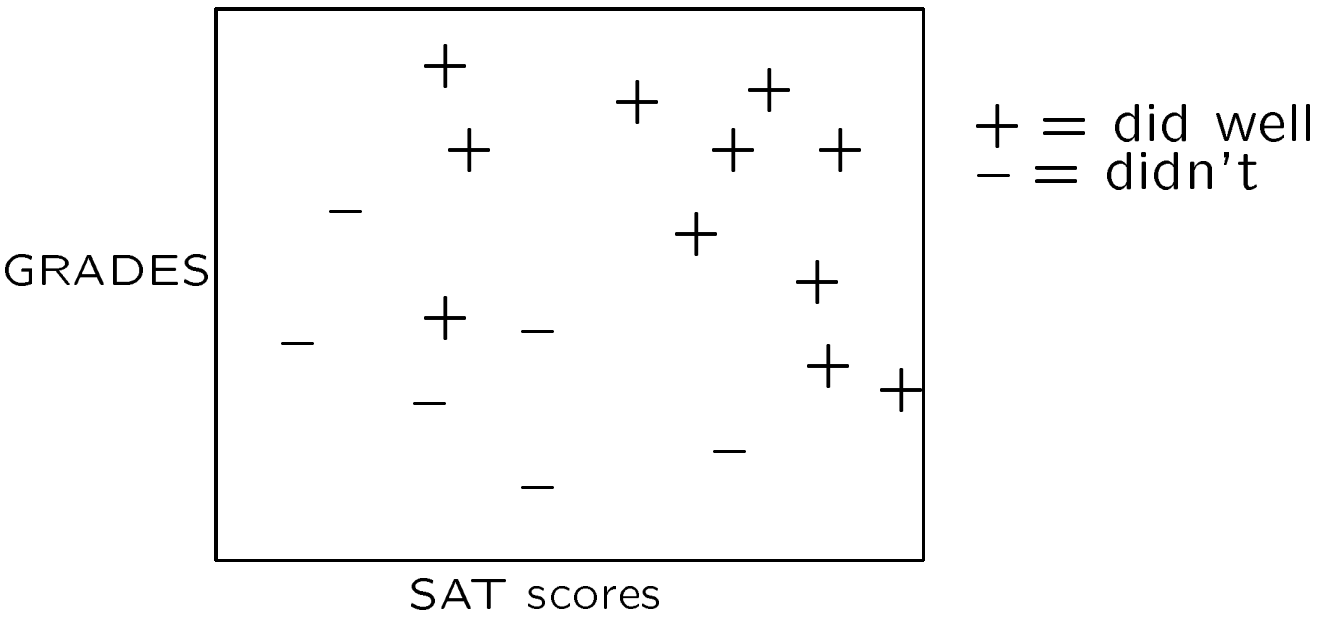
# Linear threshold functions/Perceptrons

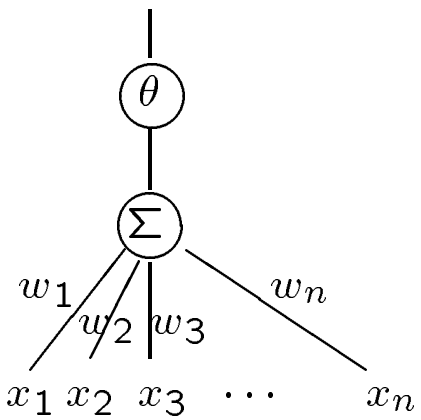



Linear threshold function: If

$$w_1 x_1 + w_2 x_2 + \ldots + w_n x_n \geq \theta$$

then output "positive", else output "negative".

View as "neuron" / "perceptron":



## Perceptron Training Procedure

For simplicity, use threshold = 0, have "dummy" input $x_0$ always 1. So,

$$\text{positive} \Leftrightarrow w_0x_0 + w_1x_1 + \ldots + w_nx_n \geq 0.$$

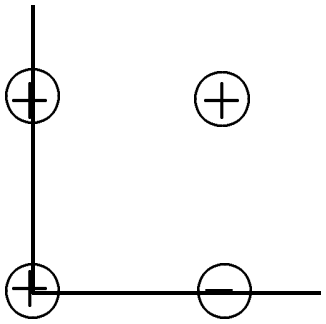
Start with some weights and cycle through data until converge or not improving:

- Mistake on positive: $\vec{w} \leftarrow \vec{w} + \vec{x}$.
- Mistake on negative: $\vec{w} \leftarrow \vec{w} - \vec{x}$.

If some examples have really large feature values, usually normalize by dividing by "length" of example:

$$\vec{w} \leftarrow \vec{w} \pm \vec{x}/|\vec{x}|$$

# Example:



**Guarantee:** If there is a consistent LTF, this will converge.

**Proof:** (following Minsky & Papert)

First simplify problem:

- Assume plane does not pass through any examples.
- Replace negative example at $(x_0, x_1, \ldots, x_n)$ by positive at $(-x_0, -x_1, \ldots, -x_n)$. So, now only positives.
- Normalize so each example $\vec{x}$ has $|\vec{x}| = 1$.

Define unit vector $w^*$ to be a correct solution. So, for each example $\vec{x}$, $w^* \cdot \vec{x} > 0$.

Define $\delta = \min(w^* \cdot \vec{x})$ (min over the data set)

Will show: sufficient decrease in angle between current weight vector and $w^*$ with each update.

## Proof, continued

Define $w$ to be current weight vector.
(Remember, unit vector $w^*$ is correct solution)

Cosine of angle between $w$ and $w^*$ is: $\frac{w \cdot w^*}{|w|}$.

Add $\vec{x}$:

$$w \cdot w^* \rightarrow w^* \cdot (w + \vec{x}) = w^* \cdot w + w^* \cdot \vec{x}.$$

(increases by at least $\delta$).

$$|w|^2 \rightarrow |w + \vec{x}|^2 = w \cdot w + x \cdot x + 2(w \cdot x).$$

(increases by at most 1)

If start with $w = \vec{0}$, in $m$ steps: denominator $\leq \sqrt{m}$ and numerator $\geq m\delta$.

So, number of steps at most $1/\delta^2$.

## Bad case for PTP:

Alternating decision list.

$(x_n \rightarrow T)$ else $(x_{n-1} \rightarrow F)$ else $(x_{n-2} \rightarrow T) \ldots$ else $(x_1 \rightarrow F)$ else $T$.

Requires exponential size coeffs.

$$-x_1 + x_2 - 2x_3 + 3x_4 - 5x_5 + 8x_6 - 13x_7 + \ldots \geq 0.$$

(magnitudes follow Fibonacci sequence)