1. LINEAR THRESHOLD FUNCTIONS (LTF)

Let \( w \cdot x = \sum_{1 \leq i \leq n} w_i x_i \) (inner product between \( w \in \mathbb{R}^n \) and \( x \in \mathbb{R}^n \))

An LTF \( f : \mathbb{R}^n \to \{0, 1\} \) has the form

\[
f(x) = \begin{cases} 
1 & \text{if } w \cdot x \geq \theta \\
0 & \text{otherwise}
\end{cases}
\]

for some weight vector \( w \in \mathbb{R}^n \) and threshold \( \theta \in \mathbb{R} \)

Every disjunction is LTF, e.g. for \( x \in \{0, 1\}^n \)

\[x_1 \lor x_2 \lor \overline{x}_3 \text{ true} \iff x_1 + x_2 + (1 - x_3) \geq 1 \iff x_1 + x_2 - x_3 \geq 0\]

Every 1-DL is LTF (why?)

2. Winnow1

Update weights multiplicatively

Learn \( k \)-sparse (i.e. involves \( k \) literals) monotone disjunctions using LTF hypothesis

\( O(k \log n) \) mistakes

When \( k \) really small (e.g. 5) and \( n \) really big, \( O(k \log n) \) is better than \( n \) (in Elimination Algorithm)

\[w_1 = \cdots = w_n = 1, \quad \theta \text{ fixed to be } n\]

On input \( x \), output hypothesis \( h(x) = 1(w \cdot x \geq \theta) \) and get \( c(x) \)

False positive \((h(x) = 1, c(x) = 0)\): For every \( i \) s.t. \( x_i = 1 \)

Set \( w_i = 0 \) (demotion, in fact elimination)

False negative \((h(x) = 0, c(x) = 1)\): For every \( i \) s.t. \( x_i = 1 \)

Double \( w_i \) (promotion)

In fact non-zero \( w_i \) is always 1, 2, 4, 8, \ldots (power of 2)

Observation: no \( w_i \) is ever negative

Observation: in every promotion step, some \( x_i \) in \( c \) has its \( w_i \) doubled

Claim: Each \( w_i \) always \(< 2n \)

Reason: When \( w_i \) is doubled, \( x_i \) must be 1 and \( w \cdot x < n \)

Claim: \#promotion steps \( \leq k \log(2n) \)

Reason: No \( x_i \) in \( c \) is ever eliminated, and is promoted \( \leq \log(2n) \) times \((k \text{ many such } x_i)\)

**Lemma 2.1.** \#elimination steps \( \leq \#promotion steps + 1 \)

**Proof.** Let \( W = \text{total weight} = \sum_{1 \leq i \leq n} w_i \) (initially \( n \))

Each elimination step \( W \) decreases by \( w \cdot x \geq n \) \((w_i \text{ becomes 0 iff } x_i = 1)\)

Each promotion step \( W \) increases by \( w \cdot x < n \) \((w_i \text{ doubled iff } x_i = 1)\)

After \( e \) elimination steps and \( p \) promotion steps, \( 0 \leq W \leq n - en + pn \), so \( e \leq p + 1 \). \( \square \)
Winnow1 makes \( \leq 2k \log(2n) + 1 = O(k \log n) \) mistakes on \( k \)-sparse monotone disjunction

**Variation:** During promotion, instead of doubling \( w_i \), can multiply \( w_i \) with \( \alpha > 1 \); Threshold \( \theta \) need not be \( n \); See Littlestone if interested

<table>
<thead>
<tr>
<th>Can Winnow1 learn non-monotone disjunction?</th>
<th>(False positive kills it e.g. ( c(x) = x_1, x^1 = 11 ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Or LTF with nonnegative weights?</td>
<td>(Not without new ideas such as Winnow2)</td>
</tr>
</tbody>
</table>

3. **Winnow2**

Can assume threshold \( \theta = 1 \) (by rescaling \( w \))

An LTF \( x \in \{0, 1\}^n \mapsto 1(w \cdot x \geq 1) \) is \( \delta \)-separated if

\[
\forall x \in \{0, 1\}^n, \text{ either } w \cdot x \geq 1 \text{ or } w \cdot x \leq 1 - \delta
\]

e.g. \( r \)-out-of-\( k \) threshold function

\[
c(x) = 1(x_{i_1} + \cdots + x_{i_k} \geq r) = 1 \left( \frac{1}{r} x_{i_1} + \cdots + \frac{1}{r} x_{i_k} \geq 1 \right)
\]

is \( 1/r \)-separated

**Winnow2**

<table>
<thead>
<tr>
<th>Initialize: ( w_1 = \cdots = w_n = 1 ), ( \theta ) fixed to be ( n ), ( \alpha ) fixed to be ( 1 + \delta/2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>On input ( x ), output hypothesis ( h(x) = 1(w \cdot x \geq \theta) ) and get ( c(x) )</td>
</tr>
<tr>
<td>False positive (( h(x) = 1, c(x) = 0 )): For every ( i ) s.t. ( x_i = 1 )</td>
</tr>
<tr>
<td>Divide ( w_i ) by ( \alpha ) (demotion)</td>
</tr>
<tr>
<td>False negative (( h(x) = 0, c(x) = 1 )): For every ( i ) s.t. ( x_i = 1 )</td>
</tr>
<tr>
<td>Multiply ( w_i ) by ( \alpha ) (promotion)</td>
</tr>
</tbody>
</table>

**Claim 3.1.** Winnow2 can learn \( \delta \)-separated LTF with nonnegative weights \( w \in \mathbb{R}^n \) with \( O((\log n)\delta^{-2} \sum_{1 \leq i \leq n} w_i) \) mistakes

Proof in Littlestone §5

\( k \)-sparse monotone disjunctions are 1-out-of-\( k \) threshold functions

Winnow2 learns \( k \)-sparse monotone disjunctions with \( O(k \log n) \) mistakes (direct proof in Blum §3.2)