Deterministic label: The target \( f \) is deterministic.

What is the prob. of seeing a tasty apple and has feature \( x_1 \)?

\[
P(X=x_1, Y=\text{yes}) = P(X=x_1) \cdot P(Y=\text{yes} \mid X=x_1)
\]

\[
0.25 \cdot 0.9 = 0.225
\]

\( x_3 = (B, \text{red}, \text{medium}, \text{crumby}) \)

\[
P(X=x_3) = 0.3
\]

Prob of seeing apple \( x_3 \) that was also tasty = 0.2

\[
P_r (Y=\text{yes} \mid X=x_3) = \frac{P_r (X=x_3, Y=\text{yes})}{P_r (X=x_3)} = \frac{0.2}{0.3} = \frac{2}{3}
\]

\[
P_r \left( S = \{ (x_1, y_1) \} \right) = \cancel{P(X=x_1, Y=y_1)}
\]

\[
P_r \left( S = \{ (x_1, y_1), (x_2, y_2) \} \right) = P(x_1, y_1) \cdot P(x_2, y_2 \mid x_1, y_1)
\]

\[
\text{(independence)}
\]

\[
(1, 1.11)
\]
\[ P_i \left( S = \{ (x_1, y_1), \ldots, (x_m, y_m) \} \right) = \prod_{i=1}^{m} P(X = x_i, Y = y_i). \]

\[ \mathbb{E}_{(x,y) \sim \mathbb{P}} \left[ \Delta(h(x), y) \right] = \mathbb{P}_Y \left[ h(x) \neq y \right] = \sum_{y \in Y} \Delta(h(x), y) \cdot P(x = x_i | Y = y). \]

\[ \text{err} \left( h_{\text{train}} \right) = 0 \quad \text{err} \left( h_{\text{train}} \right) = \frac{1}{2} \quad \text{large}. \]

Overfitting:
* S too small.
* \( X \) is rich, \( H \) is rich.
* Memoize \( S \) without "learning" patterns.

DTS are expressive.
H : 2

H : AND OR Decision... 

Assumption:
- Distances
- Features meaningful
- Assumption on H

<table>
<thead>
<tr>
<th>H</th>
<th>16</th>
</tr>
</thead>
</table>
| 2 | 2 | all functions

Occam's Razor:

**Good:** World is simple, structured.
Simple → fitted.

**Bad:** World is not simple.
Simplicity perpetuating stereotypes/hoaxes.