GENETIC ALGORITHMS

October 31: Lecture 15
Biological Evolution

**Charles Darwin:** Evolution by natural selection
‘The survival of the fittest’

**Evolutionary Computing:** An attempt to mimic the principles of biological evolution in the context of machine learning

**More concretely:** Attempt to learn a ‘fit hypothesis’
What is a fit hypothesis?

How do we define fitness?

Between two hypotheses for the same target concept which one should we consider more fit? What can we use?

In general we can perform our selection based on a **fitness function**.
Biological Evolution and Randomness

Evolution is not deterministic.

Imagine a very fit person born in a car rushing to the hospital.....

Or a very lucky unfit person.

In our context: Apply randomness in selection. Randomness gets smoothed out by population.
Biological Evolution and String Encodings

**DNA:** The genetic **code** of every known species

The DNA is a sequence of letters from an alphabet of 4. They collectively encode information.

**In our context:** We will **encode** (in other words represent) hypotheses with binary strings.
Genetic Operators

- Point Mutation

\[
\begin{array}{c}
11101001000 \\
\rightarrow
\end{array}
\begin{array}{c}
11101011000
\end{array}
\]

- Single Point Crossover

\[
\begin{array}{c}
11101001000 \\
\downarrow
\end{array}
\begin{array}{c}
11111000000
\end{array}
\begin{array}{c}
11101010101
\end{array}
\begin{array}{c}
0001010101
\downarrow
\end{array}
\begin{array}{c}
0001001000
\end{array}
\]

Crossover Mask
Genetic Operators

- Two-point Crossover
  
  ![Two-point Crossover Diagram]

- Uniform Crossover
  
  ![Uniform Crossover Diagram]
Representing Hypotheses

- How to represent if-then-else rules:

- Possible values:
  - **Outlook**: Sunny, Overcast, Rain
  - **Wind**: Strong, Weak

\[(\text{Outlook} = \text{Overcast} \lor \text{Rain}) \land (\text{Wind} = \text{Strong})\]

Representation: 011 10
Representing Hypotheses

- Crossover adjusted to accommodate different lengths

- From an actual example:
  
  \[
  \text{IF } a_1 = T \land a_2 = F \text{ THEN } c = T; \quad \text{IF } a_2 = T \text{ THEN } c = F
  \]

- Representation String:

  \[
  a_1 \quad a_2 \quad c \quad a_1 \quad a_2 \quad c
  \]

  \[
  10 \quad 01 \quad 1 \quad 11 \quad 10 \quad 0
  \]
Representing Hypotheses

$h_1: \begin{array}{ccc} a_1 & a_2 & c \\ 10 & 01 & 1 \end{array} \quad h_2: \begin{array}{ccc} a_1 & a_2 & c \\ 01 & 11 & 0 \end{array}

Random positions

Distances from beginning of rule: 1 and 3 respectively

Three possible positions of crossover in second parent

One of the three is selected at random
Representing Hypotheses

• Crossover adjusted to accommodate different lengths

• If the points (1,3) are chosen as crossover, here is what happens:

\[
\begin{align*}
  h_1 : & \quad a_1 \quad a_2 \quad c \\
  & \quad 10 \quad 01 \quad 1 \\
  h_2 : & \quad a_1 \quad a_2 \quad c \\
  & \quad 01 \quad 11 \quad 0 \\
  h_3 : & \quad a_1 \quad a_2 \quad c \\
  & \quad 11 \quad 10 \quad 0 \\
  h_4 : & \quad a_1 \quad a_2 \quad c \\
  & \quad 00 \quad 01 \quad 1 \\
\end{align*}
\]

h3 is produced by replacing the (1,8) segment of h1 by the (1,3) segment of h2

h4 is produced by replacing the (1,3) segment of h2 by the (1,8) segment of h1
A prototypical algorithm

\[
\text{GA}(\text{Fitness, Fitness\_threshold, p, r, m})
\]

- **Fitness**: A function that assigns an evaluation score, given a hypothesis.
- **Fitness\_threshold**: A threshold specifying the termination criterion.
- **p**: The number of hypotheses to be included in the population.
- **r**: The fraction of the population to be replaced by Crossover at each step.
- **m**: The mutation rate.

**Initialization**

- **Initialize population**: \( P \leftarrow \text{Generate } p \text{ hypotheses at random} \)
- **Evaluate**: For each \( h \) in \( P \), compute \( \text{Fitness}(h) \)
A prototypical algorithm

Main Loop

- While \[
  \left[ \max_h \operatorname{Fitness}(h) \right] < \operatorname{Fitness\_threshold}
\] do

  Create a new generation, \( P_S \):
  1. Select: Probabilistically select \((1 - r)p\) members of \( P \) to add to \( P_S \). The probability \( \Pr(h_i) \) of selecting hypothesis \( h_i \) from \( P \) is given by

     \[
     \Pr(h_i) = \frac{\operatorname{Fitness}(h_i)}{\sum_{j=1}^{p} \operatorname{Fitness}(h_j)}
     \]

  2. Crossover: Probabilistically select \( \frac{r}{2}p \) pairs of hypotheses from \( P \), according to \( \Pr(h_i) \) given above. For each pair, \( (h_1, h_2) \), produce two offspring by applying the Crossover operator. Add all offspring to \( P_S \).

  3. Mutate: Choose \( m \) percent of the members of \( P_S \) with uniform probability. For each, invert one randomly selected bit in its representation.

  4. Update: \( P \leftarrow P_S \).

  5. Evaluate: for each \( h \) in \( P \), compute \( \operatorname{Fitness}(h) \).
Advantages and Disadvantages

• The procedure can jump suddenly to other populations. No local maxima.

• Crowding