Previous Lecture

- Genetic Programming
  - Evolutionary algorithms with tree representation
    - Tree-crossover, tree-mutation, tree-generation
    - Ideal for some applications
    - Bloat
  - Automatic generation of programs
    - Tree-, list-, network-based representations
Case Study: Galaxy Brightness

- Astrophysics application
  - Problem: find a formula for the radial brightness of a typical galaxies

Negative images of typical elliptical and spiral galaxy
Traditional Approach

- Guesswork...
  - Hubble's law:
    - Physically inspired
    - Bad fit...
  - De Vaucouleurs law:
    - No physical interpretation
    - Reasonable fit
  - Come up with a formula
    - Use square error fitting to adjust parameters
      - (such as 3.33 above)

\[
I(r) = \frac{I_0}{(\frac{r}{a} + 1)^2}
\]

\[
I(r) = I_c e^{-3.33 \left( \left( \frac{r}{r_c} \right)^{1/4} - 1 \right)}
\]
Typical observed profiles

- 18 elliptical galaxies
  - Coma cluster, near-infrared
Problems with current approach

- Lots of new data coming in
  - Great variety of shapes
  - Generalized model?
    - Bulges, shallower vs. steeper profiles
  - Physical interpretation?
Evolutionary Approach

- Suggestions?
Representation
Fitness Function

- Hits
  - Number of function points within tolerances
    - (max: 50 points)
    - Depends on tolerances
  - Size penalty
    - Simpler functions are better
    - $F = \text{hits} - w \times \text{nodes}$
# GP details

<table>
<thead>
<tr>
<th>Target</th>
<th>To find a regression mathematical function based on the galaxy profile given</th>
</tr>
</thead>
<tbody>
<tr>
<td>Terminals</td>
<td>( R, R = [\overline{-10, 10}] )</td>
</tr>
<tr>
<td>Non-terminals</td>
<td>( +, -, *, /, \exp, \sin, \cos, \log )</td>
</tr>
<tr>
<td>Crossover rate</td>
<td>0.9</td>
</tr>
<tr>
<td>Mutation rate</td>
<td>0.01</td>
</tr>
<tr>
<td>Population size</td>
<td>6,000</td>
</tr>
<tr>
<td>Maximum no. of generations</td>
<td>100</td>
</tr>
<tr>
<td>Termination criterion</td>
<td>Generation limit or Time limit, whichever reached first</td>
</tr>
<tr>
<td>Selection strategy</td>
<td>Tournament selection, Size = 6</td>
</tr>
<tr>
<td>Max depth of individual expressions</td>
<td>17</td>
</tr>
<tr>
<td>Hit criterion</td>
<td>0.005</td>
</tr>
<tr>
<td>Fitness criterion</td>
<td>Hits (-0.01 \times) the length of the expression</td>
</tr>
<tr>
<td>Max depth of initial individual expressions</td>
<td>6</td>
</tr>
<tr>
<td>Maximum run time (hours)</td>
<td>6</td>
</tr>
</tbody>
</table>
Results

- 18 models
  - All with fitness 50
  - Different tree structures
- What now?
Manual Simplification

- Simplify trees
  - Get rid of unused branches
  - Transform into most compact form
- Parameterize
  - Turn real-valued terminals into parameters

\[ f_{g1} = \frac{a + b}{c + r} \]
\[ f_{g2} = a' + \frac{b'}{b' + c' \cdot r^2} \]
Evolutionary Parameter Optimization

- Representation?
- Crossover, Mutation?
- Fitness?
## Results

<table>
<thead>
<tr>
<th>Profiles</th>
<th>$\chi^2$ for $f_{g1}$ Using hit criteria = 0.005</th>
<th>$\chi^2$ for $f_{g1}$ Using hit criteria = 0.002</th>
<th>$\chi^2$ for $f_{g2}$ Using hit criteria = 0.005</th>
<th>$\chi^2$ for $f_{g2}$ Using hit criteria = 0.002</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7.4460</td>
<td>7.8961</td>
<td>11.4012</td>
<td>1.1141</td>
</tr>
<tr>
<td>2</td>
<td>9.1809</td>
<td>6.2876</td>
<td>12.8405</td>
<td>1.8633</td>
</tr>
<tr>
<td>3</td>
<td>8.7796</td>
<td>5.0199</td>
<td>6.3452</td>
<td>0.7113</td>
</tr>
<tr>
<td>4</td>
<td>6.8782</td>
<td>6.8221</td>
<td>2.8035</td>
<td>2.8385</td>
</tr>
<tr>
<td>5</td>
<td>4.1948</td>
<td>3.8105</td>
<td>5.7141</td>
<td>0.8685</td>
</tr>
<tr>
<td>6</td>
<td>3.1867</td>
<td>2.7282</td>
<td>3.2442</td>
<td>0.7194</td>
</tr>
<tr>
<td>7</td>
<td>9.7797</td>
<td>4.8050</td>
<td>17.9798</td>
<td>0.9289</td>
</tr>
<tr>
<td>8</td>
<td>6.8756</td>
<td>4.9609</td>
<td>11.0505</td>
<td>0.1353</td>
</tr>
<tr>
<td>9</td>
<td>5.8373</td>
<td>5.0018</td>
<td>3.4484</td>
<td>1.0390</td>
</tr>
<tr>
<td>10</td>
<td>2.8263</td>
<td>2.8807</td>
<td>1.9078</td>
<td>0.3318</td>
</tr>
<tr>
<td>11</td>
<td>34.0802</td>
<td>8.2256</td>
<td>25.2975</td>
<td>7.9733</td>
</tr>
<tr>
<td>12</td>
<td>5.8477</td>
<td>5.0861</td>
<td>7.9886</td>
<td>1.5150</td>
</tr>
<tr>
<td>13</td>
<td>7.6026</td>
<td>7.6528</td>
<td>10.6681</td>
<td>1.2615</td>
</tr>
<tr>
<td>14</td>
<td>9.1534</td>
<td>5.1829</td>
<td>13.8083</td>
<td>1.2896</td>
</tr>
<tr>
<td>15</td>
<td>5.9548</td>
<td>5.8170</td>
<td>2.1233</td>
<td>1.2062</td>
</tr>
<tr>
<td>16</td>
<td>4.8048</td>
<td>1.98719</td>
<td>3.7705</td>
<td>0.6759</td>
</tr>
<tr>
<td>17</td>
<td>2.9778</td>
<td>3.0106</td>
<td>1.3418</td>
<td>0.3138</td>
</tr>
<tr>
<td>18</td>
<td>1.2161</td>
<td>1.0233</td>
<td>0.8031</td>
<td>0.0796</td>
</tr>
</tbody>
</table>

$\chi^2 < 2.0 : good\ fit$
Critique

- Functions not evolved for generalization
  - Evolving functions for each profile, then fitting to other functions.
- Evolve generalized functions in one step?
  - Parameter fitting as sub-algorithm?
- Mini project?
• Awards for Human-Competitive results produced by genetic and evolutionary computation
• $5000, $3000, $3000 awards
Summary

• Treat EA's as 'construction toys'
  – Pick representation suitable to the problem
  – Pick operators suitable to representation
  – Pick population, selection, etc suitable for the problem
  – Write custom fitness function
  – Add speciation, constraint handling, etc. as required

• Human-competitive result