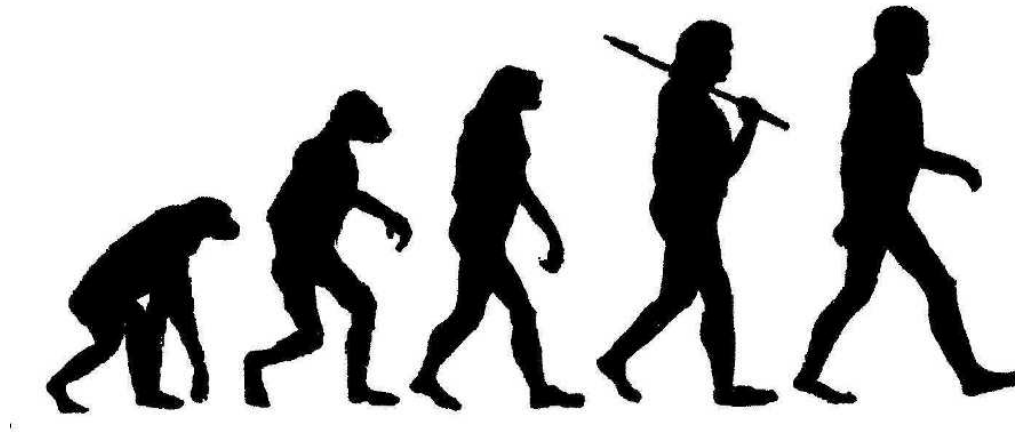Pattern Recognition II

Evolutionary Feature Selection

Thomas Bergmüller, Lefteris Christopoulous and Martin Schnöll
4th November 2014

Evolutionary Algorithms for Feature Selection
Introduction

- Evolutionary Algorithms (EA), used for feature selection, belong to the **parallel feature selection techniques** (a complete feature subset is generated at once)

- They are often used for nonlinear, high-dimensional problems of exponential complexity

Evolutionary Algorithms for Feature Selection
Basic procedure

1. Initialize Population of size N

2. Select 2 parents for mating (based on some fitness)

3. Mating (Generation of 2 new children)

4. Repeat steps 2 & 3 till initial population size N is reached -> this forms a new generation

Evolutionary Algorithms for Feature Selection
Initialization

- How are the initial chromosomes (feature subsets) generated?

- By **Permutation Encoding**

- **Example (for 5 features and a given subset size of 2):**

  3 5 1 4 2 -> **3 5** 1 4 2 -> 0 0 1 0 1

  1 2 4 3 5 -> **1 2** 4 3 5 -> 1 1 0 0 0

Evolutionary Algorithms for Feature Selection
Selection of parents & mating

- The parents for mating are selected based on a fitness function (in our case the accuracy of the kNN-classifier)

- How are the children generated? How do they look like?

- **Crossover (partially matched crossover, $p_c = 0.6$):**

  3|5 1|4 2 -> **3|2** 4|4 2 -> **3|2** 4|1 5

  1|2 4|3 5 -> **1|5** 1|3 5 -> **4|5** 1|3 2

- **Mutation ($p_m = 0.01$):**

  3 **2** 4 **1** 5 -> 3 **1** 4 **2** 5

  4 **5** 1 **3** 2 -> 4 **3** 1 **5** 2

- If no crossover or mutation happens, the children are exact copies of the parents

Evolutionary Algorithms for Feature Selection
Parameters & Implementation

- Summary of arbitrary parameters (with default values):
  - Population Size = 50
  - Generations = 100
  - Crossover Probability p_c = 0.6
  - Mutation Probability p_m = 0.01

- Other parameters to consider
  - K for kNN classifier
  - How many runs? (Due to random initialization, more runs and averaging should probably be implemented)

Evolutionary Algorithms for Feature Selection
Project Software and Milestones

- For the implementation we will use the **JEvolution package (Java),** which was developed at University of Salzburg

- Milestones
  - Testing of JEvolution package (End of November)
  - Actual implementation & computation (December)
  - Evaluation, presentation & documentation (January)