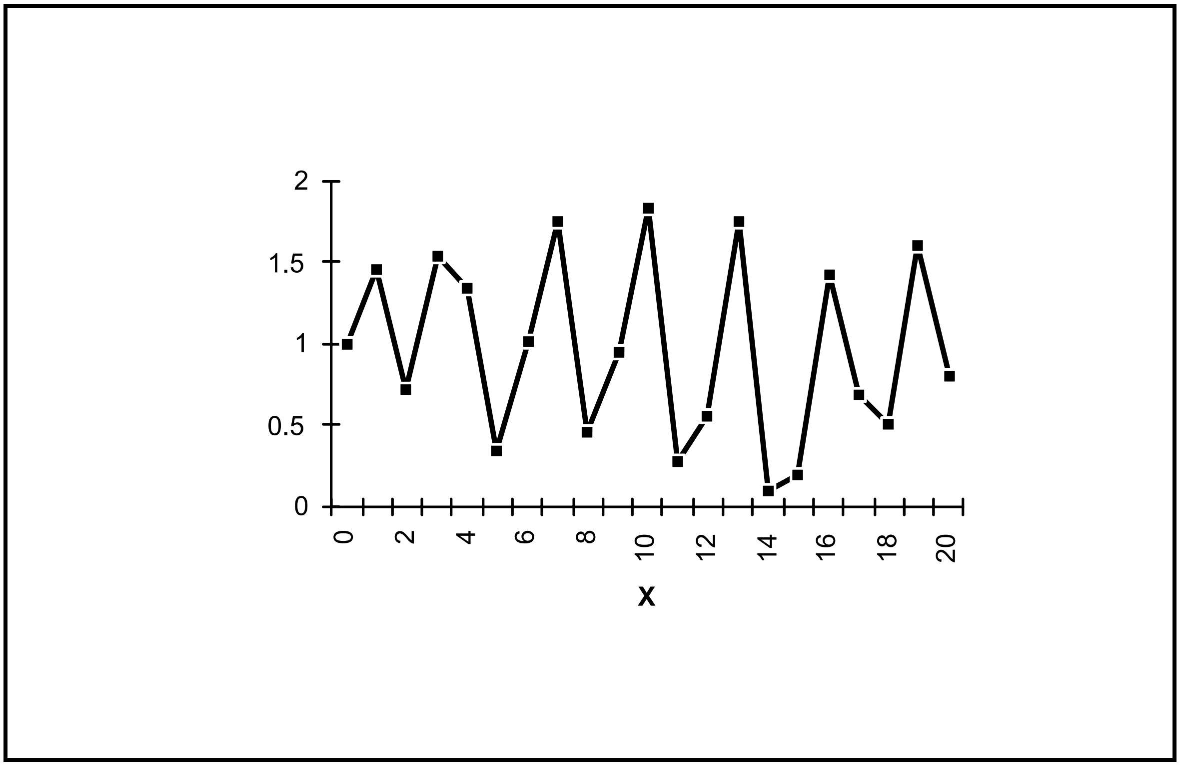
COMPARISON OF SEARCH AND LEARNING METHODS

SIMULATED ANNEALING



FIVE MAJOR PREPARATORY STEPS FOR SIMULATED ANNEALING

- **(1) determining the representation scheme which maps the individual points in the search space of the problem into a structure**
- **(2) determining the operation for modifying the structures (including the neighborhood and possibly the step size)**
- **(3) determining the energy (fitness) function**
- **(4) determining the parameters**
  - **notably, the annealing schedule**
  - **number of possible neighbors that can be produced by the modifying operation**
  - **(often) a step size for controlling the distance in the search space between the current structure and possible neighbors that will be produced by the modifying operation**
- **(5) determining the criterion for terminating a run**

SIMULATED ANNEALING

- **Kirkpatrick, Gelatt, and Vecchi (1983)**
- **user-defined domain-specific structure**
- **user-defined method for modifying a structure**
- **Metropolis algorithm**
- **Single point in search space is modified into another single point**
- **Either a random initial structure or an initial structure that is already known to be fairly good**
- **Zero-is-best energy function (minimizing)**
- **Terminate when no move from the current structure is an improvement and the annealing schedule has been completely executed**

STEPS OF SIMULATED ANNEALING

- **A domain-specific method for modifying any existing structure is defined by the user.**

- **The result of the modification can be one of several neighboring structures in the search space of possible structures.**

- **The existing structure is tentatively modified using the method of modification and its energy level is determined.**

STEPS OF SIMULATED ANNEALING

• **The Metropolis algorithm is applied. If the energy level of the modification is an improvement, the modification is always accepted. If the energy level of the modification is not an improvement, the modification may still be accepted with a certain probability determined by the Boltzmann equation. This probability of acceptance is greater if the energy difference is small and it is greater if the temperature parameter *T* is high.**

PROBABILITY OF ACCEPTING A CANDIDATE NEW POINT

- **if $\Delta E = E_{new} - E_{current} < 0$, the probability of accepting candidate point is 1.0**

- **if $\Delta E = E_{new} - E_{current} > 0$, the probability of accepting candidate point is**

$$e^{-\Delta E / kT} = e^{-(E_{new} - E_{current}) / kT}$$

- **Exponential allocation of trials**

FOUR MAJOR PREPARATORY STEPS FOR THE CONVENTIONAL GENETIC ALGORITHM OPERATING ON FIXED-LENGTH STRINGS

(1) the representation scheme (i.e., the alphabet size *K*, the chromosome length *L*, and the mapping between the problem and the chromosome),
(2) the fitness measure,
(3) the parameters and qualitative variables for controlling the algorithm, and
(4) the criterion for terminating a run and the method for designating the result.

FOUR PREPARATORY STEPS FOR HILL CLIMBING

(1) the representation scheme (i.e., the mapping between the problem and the points in the multi-dimensional search space),
(2) the fitness measure for points in the search space,
(3) the parameters (i.e., the step size and the number of alternative points to be considered before a step is taken) for controlling the algorithm, and
(4) the criterion for terminating a run and the method for designating the result.

FOUR PREPARATORY STEPS FOR THE *EVOLUTIONSSTRATEGIE* (ES)

(1) the representation scheme (i.e., the number of components of the real-valued vector and the mapping between the problem and the components of the vector),
(2) the fitness (payoff) measure,
(3) the parameters for controlling the algorithm, and
(4) the criterion for terminating a run and the method for designating the result.

TEN PREPARATORY STEPS FOR A NON-RECURRENT NEURAL NETWORK

(1) the architecture of the network (e.g., number of layers, number of processing elements in each layer),
(2) the connectivity of the network (e.g., full or partial connectivity between consecutive layers; whether or not the network is recurrent; what connections from one layer to earlier layers are permitted),
(3) the type of processing element used (e.g., linear threshold processing element, sigmoid processing element),
(4) the training paradigm (e.g., back propagation),
(5) the inputs to the network,

TEN PREPARATORY STEPS FOR A NON-RECURRENT NEURAL NETWORK – CONTINUED

(6) the outputs of the network,
(7) the training cases to be used,
(8) the error measure,
(9) the values of the numerical parameters for controlling the run (i.e., learning rate for back propagation, average magnitude of initial random weights, etc.), and
(10) the criterion for designating a result and terminating a run (e.g., the criterion for stopping training).

SIX PREPARATORY STEPS FOR INDUCTION OF A DECISION TREE USING ID3

(1) the set of class names,
(2) the set of attribute-testing functions,
(3) the heuristic entropy-based fitness measure to be used,
(4) the examples (training cases) to be used,
(5) the values of the numerical parameters (e.g., branching factor) for controlling the run, and
(6) the criterion for designating a result and terminating a run.

FIVE MAJOR PREPARATORY STEPS FOR SIMULATED ANNEALING

- **(1) determining the representation scheme which maps the individual points in the search space of the problem into a structure**
- **(2) determining the operation for modifying the structures (including the neighborhood and possibly the step size)**
- **(3) determining the energy (fitness) function**
- **(4) determining the parameters**
  - **notably, the annealing schedule**
  - **number of possible neighbors that can be produced by the modifying operation**
  - **(often) a step size for controlling the distance in the search space between the current structure and possible neighbors that will be produced by the modifying operation**
- **(5) determining the criterion for terminating a run**

ADAPTIVE SYSTEMS

- **the structures that undergo adaptation,**
- **the initial structures,**
- **the fitness measure that evaluates the structures,**
- **the operations that modify the structures,**
- **the state (memory) of the system at each stage,**
- **the method for terminating the process,**
- **the method for designating a result, and**
- **the parameters that control the process.**

COMPARISON OF STRUCTURES UNDERGOING ADAPTATION FOR SEVERAL ADAPTIVE OR LEARNING PARADIGMS

Paradigm	**Structure undergoing adaptation**
Genetic programming:	**Population consisting of hierarchical compositions of functions from the function set and terminals from the terminal set.**
Genetic algorithm:	**Population consisting of fixed-length character strings.**
Hill climbing:	**A single point in the search space.**
Evolutionsstrategie:	**A single point (real-valued vector) in the search space of such vectors.**
Neural network:	**A single vector of weights in weight space.**
Decision tree:	**A single rooted, point-labeled, line-labeled decision tree in the space of possible decision trees.**
Simulated annealing:	**A single domain specific structure in the search space.**

COMPARISON OF INITIAL STRUCTURES FOR SEVERAL ADAPTIVE OR LEARNING PARADIGMS

Paradigm	Initial Structures
Genetic programming:	**A population of randomly created hierarchical compositions of functions from the function set and terminals from the terminal set.**
Genetic algorithm:	**A population of randomly created fixed-length character strings over the given alphabet.**
Hill climbing:	**Usually a random initial point in the search space, but possibly a point believed by the user to be a good starting point for the search.**
Evolutionsstrategie:	**Usually a random initial point (real-valued vector) in the search space, but possibly a point believed by the user to be a good starting point for the search.**
Neural network:	**For backpropagation, a randomly created initial weight vector consisting of small weights.**
Decision tree:	**A decision tree consisting of one internal point (i.e., the root) labeled with the single attribute testing function that maximizes the payoff measure.**
Simulated annealing:	**Either a random initial structure or a structure that already performs fairly well.**

COMPARISON OF FITNESS MEASURES FOR SEVERAL ADAPTIVE OR LEARNING PARADIGMS

Paradigm	**Fitness measure**
Genetic programming:	**Normalized fitness.**
Genetic algorithm:	**Normalized fitness.**
Hill climbing:	**Fitness (payoff) of a point in the search space.**
Evolutionsstrategie:	**Fitness (payoff) of a point (real-valued vector) in the search space.**
Neural network:	**Sum, taken over a number of training examples, of the square of errors between the output signal produced by neural network and the desired output signal.**
Decision tree:	**Entropy of classification performed by the partially constructed decision tree.**
Simulated annealing:	**Energy of the current structure.**

COMPARISON OF THE OPERATIONS FOR MODIFYING THE STRUCTURES FOR SEVERAL ADAPTIVE OR LEARNING PARADIGMS

Paradigm	Operations for modifying the structures
Genetic programming:	**Reproduction and crossover.**
Genetic algorithm:	**Reproduction, crossover, and occasional mutation**
Hill climbing:	**Use gradient information to move away from the current point to the best of the tested nearby points (i.e., move in the direction of steepest slope in the improving direction).**
Evolutionsstrategie:	**The Gaussian mutation operation mutates the one current point (real-valued vector) in the search space to provide a tentative offspring by adding zero-mean normally distributed random numbers to the components of the parental vector. The parent and offspring are compared as to fitness (payoff) and the better of the two is chosen (provided that the offspring does not violate any of the problem specific constraints as to its structure).**
Neural network:	**Modify the weights in the weight vector using the error measure and the Delta rule.**

Decision tree:	**For every point in a partially constructed decision tree, entropy is evaluated for every possible way of adding one internal point containing an attribute testing function or by labeling the current point with a class name. The alternative that maximizes the entropy measure is chosen.**
Simulated annealing:	**A domain-specific method for modifying any existing structure is defined by the user. The result of the modification can be one of several neighboring structures in the search space of possible structures. The existing structure is tentatively modified using the method of modification and its energy level is determined. Then the Metropolis algorithm is applied. If the energy level of the modification is an improvement, the modification is always accepted. If the energy level of the modification is not an improvement, the modification may still be accepted with a certain probability determined by the Boltzmann equation. This probability of acceptance is greater if the energy difference is small and it is greater if the temperature parameter *T* is high.**

COMPARISON OF THE STATE (MEMORY) OF THE SYSTEM FOR SEVERAL ADAPTIVE OR LEARNING PARADIGMS

Paradigm	**State (memory)**
Genetic programming:	**The population.**
Genetic algorithm:	**The population.**
Hill climbing:	**Current single point in the search space.**
Evolutionsstrategie:	**Current single point in the search space.**
Neural network:	**The current single weight vector in weight space.**
Decision tree:	**The current single partially constructed decision tree.**
Simulated annealing:	**The current single structure.**

COMPARISON OF THE TERMINATION CRITERIA FOR SEVERAL ADAPTIVE OR LEARNING PARADIGMS

Paradigm	Termination criterion
Genetic programming:	**After a specified number of generations or when some acceptable and recognizable result is obtained.**
Genetic algorithm:	**After a specified number of generations or when some acceptable and recognizable result is obtained.**
Hill climbing:	**When no tested nearby alternative point is an improvement over the current point (which may not be the global optimum).**
Evolutionsstrategie:	**After a specified number of generations or when some acceptable and recognizable result is obtained.**
Neural network:	**When no further improvement is occurring from the current point in the weight space.**
Decision tree:	**When no further improvement can occur by replacing endpoints (leaves) of the tree with additional attribute-testing functions.**
Simulated annealing:	**When no move from the current structure is an improvement and the annealing schedule has been completely executed.**

COMPARISON OF THE METHOD FOR RESULT DESIGNATION FOR SEVERAL ADAPTIVE OR LEARNING PARADIGMS

Paradigm	**Result designation**
Genetic programming:	**The best-so-far individual.**
Genetic algorithm:	**The best-so-far individual.**
Hill climbing:	**The current point in the search space at the time of termination.**
Evolutionsstrategie:	**The current point in the search space at the time of termination.**
Neural network:	**The current weight vector in the weight space at the time of termination.**
Decision tree:	**The current decision tree at the time of termination.**
Simulated annealing:	**The current structure at the time of termination.**

COMPARISON OF THE CONTROL PARAMETERS FOR SEVERAL ADAPTIVE OR LEARNING PARADIGMS

Paradigm	Control parameters

Genetic programming:	<u>Major parameters</u> **Population size *M*.** **Maximum number *G* of generations to be run.** <u>Minor parameters</u> **Crossover probability p_c.** **Reproduction probability p_r.** **Probability p_{ip} of choosing internal points for crossover.** **Maximum size D_c for S-expressions created during the run.** **Maximum size D_i for initial random S-expressions.** **Probability p_m of mutation.** **Probability p_p of permutation.** **Frequency f_{ed} of editing.** **Probability p_{en} of encapsulation.** **Condition for decimation.** **Decimation target percentage p_d.** <u>Qualitative variables</u> **Generative method for initial random population.** **Basic selection method.** **Spousal selection method.** **Adjusted fitness usage.** **Over-Selection usage.** **Elitist strategy usage.**

Genetic algorithm:	**Population size *M*.** **Maximum number *G* of generations to be run.** **Crossover probability p_c.** **Reproduction probability p_r.** **Mutation probability p_m.** **Inversion probability p_i.** **Basic selection method.** **Spousal selection method.** **Elitist strategy usage.**
Hill climbing:	**Step size.** **Number of alternative points to be considered before a step is taken.**
Evolutionsstrategie:	**Initial standard deviation for the Gaussian mutation operator.** **Step size controls for changing the standard deviation for the Gaussian mutation operator.**

Neural network:	**Number of layers in the neural network. Number of processing elements in each layer. Thresholds of the processing elements. Biases, if any, of the processing elements. Whether the network is feed-forward only or recurrent (and, if it is recurrent, what interconnectivity between layers is permitted). Map of the connectivity allowed between a processing element in one layer and other processing elements in the network. Learning rate (for back propagation). Average magnitude of initial random weights (for back propagation).**
Decision tree:	**Branching factor.**
Simulated annealing:	**Annealing schedule for varying the temperature (i.e., decreasing it) over the time steps of the process. The number of possible neighbors that can be produced by the modifying operation. Often, a step size for controlling the distance in the search space between the current structure and possible neighbors that will be produced by the modifying operation.**