Artificial Neural Networks
Lecture Notes

Stephen Lucci, PhD

Part 10

About this file:

- This is the printer-friendly version of the file "*lecture10.htm*". In case the page is not properly displayed, use IE 5 or higher.
- Since this is an offline page, each file "lecture*nn*.htm" (for instance "lecture10.htm") is accompanied with a "img*nn*" folder (for instance "img10") containing the images which make part of the notes. So, to see the images, Each html file must be kept in the same directory (folder) as its corresponding "img*nn*" folder.
- If you have trouble reading the contents of this file, or in case of transcription errors, email gi0062@bcmail.brooklyn.cuny.edu
- Acknowledgments:
Background image is from http://www.anatomy.usyd.edu.au/online/neuroanatomy/tutorial1/tutorial1.html (edited) at the University of Sydney Neuroanatomy web page. Mathematics symbols images are from metamath.org's GIF images for Math Symbols web page. Other image credits are given where noted, the remainder are native to this file.

Contents

Financial Modeling

- Predicting success for one's investments.
- In any model, the underlying component is <u>time</u>.

Converting Temporal Sequences into Spatial Patterns

- We have to convert *temporal sequences* (eg, is market up since yesterday?) into *spatial patterns*.

- In order to do this, *Discrete Time Sampling* is employed:
  A continuously variable signal is quantized at regular time intervals. A sequence of n samples is concatenated to form a single pattern that encapsulates a quantized signature of the signal.

The Dow Jones Industrial Average

- The Dow Jones Industrial Average is a single value that provides an instantaneous indication of the state of the market.

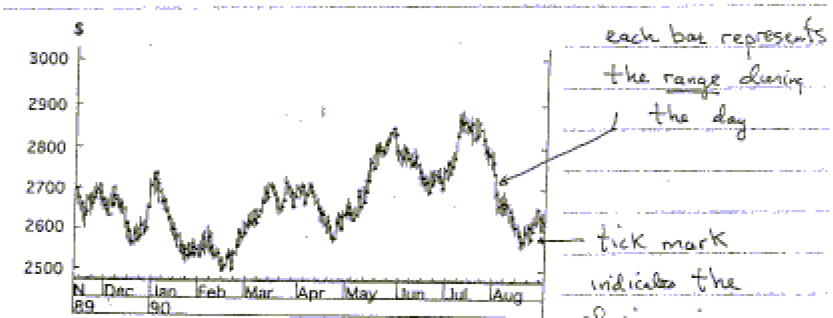


Figure 5.1 This graph shows the daily high and low of the Dow Jones Industrial Average from November 1989 through August 1990. In this chart, the vertical axis represents the numerical value of the average, while the horizontal axis depicts the passage of time. Each bar represents the range of the average during the day, while the tick mark on the range indicates the closing average. Thus, we have quantized the dynamic behavior of the market by taking a daily sample of the high, low, and closing points of the average. When looking at this chart, do you see a series of discrete points, or do you see a line moving from left to right, indicating the long-term behavior of the market?

However ...

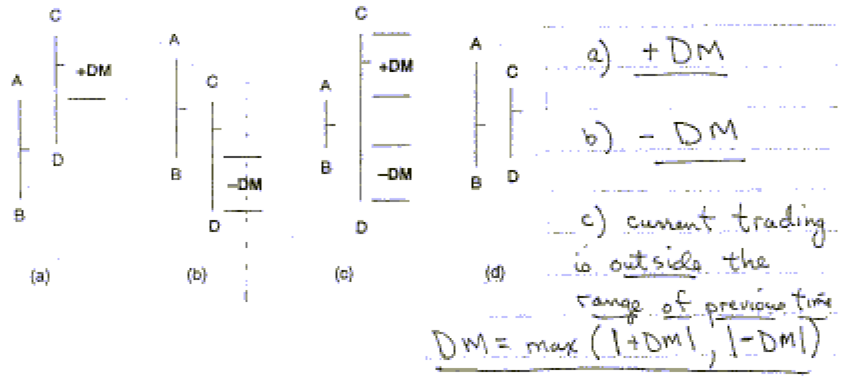
- The Dow Jones Average, by itself, offers practically no insight into what the market might do <u>tomorrow</u>.

- Gaining insight into the direction the market (or a particular stock) may be going, a finanacial analyst will consider certain <u>mathematical indicators</u>:
  - *ADX* - market-intensity indicator - is the market trending?
  - *MACD*- moving-average convergence/divergence - optimal buy & sell signals in a trending market.
  - *Slow Stochastic Analysis* - complements MACD.

Average Directional Movement

- ADX: The high and low values of the market at the current time are compared with the high and low values at the previous time.

- The difference values obtained indicate the *plus directional movement* (+DM) and the *minus directional movement* (-DM)



- In the above figure, the vertical lines represent the high and low points during a given day, and the tick on a vertical line represents the closing value for that day. We have,
  (a) +DM
  (b) -DM
  (c)Current trading is *outside* the range of previous time
  DM = MAX(|+DM| , |-DM|).
  (d) The trading range at the current time is *inside* the range of trading at previous time.
  DM=0.

- DI: The *Directional Indicator* (wilder) is the percentage of the price range that is directional for the given time period.

  DI = DM / TR

  where TR is the *True Range*.

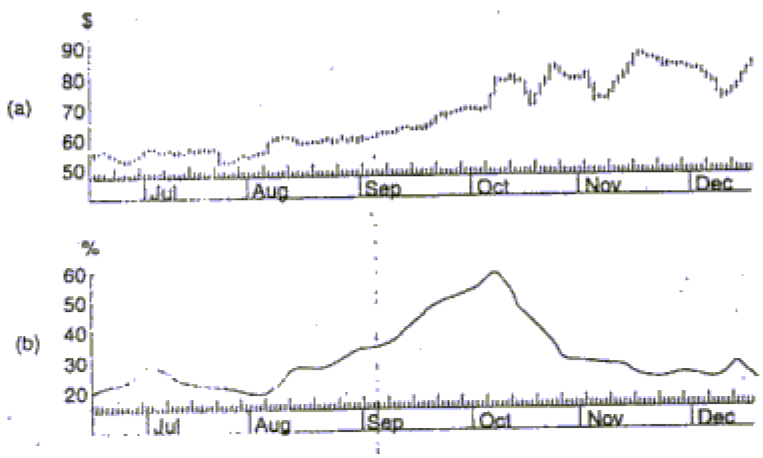
- TR is the largest of:
  - The difference between the current high and low value
  - The difference between the current high and the previous closing value
  - The difference beetween the current low and the previous closing value.

- Note: DI may be positive or negative.

- Positive Notation - wilder, defines two separate indicators
  - +DI indicates a time period with positive DI.
  - -DI absolute value of DI where DI is negative.

- ADX - *Average Directional Movement* is a *smoothed* moving average of the DI values across an interval of n time periods.

- It is useful to convert DI (Directional Indicator) to a *Directional Movement Index* (DMI) which indicates the magnitude of the trend on a scale from 0 to 100.
- The ADX is computed as an n-period moving average of the DMI.



(a) Discrete time behavior of a stock on NYSE over six months.
You can see *daily fluctuations* yet *long-term increase*. (b) ADX for the same stock.

Notice how ADX peaks concomitant with point in time when stock growth trend diminishes.

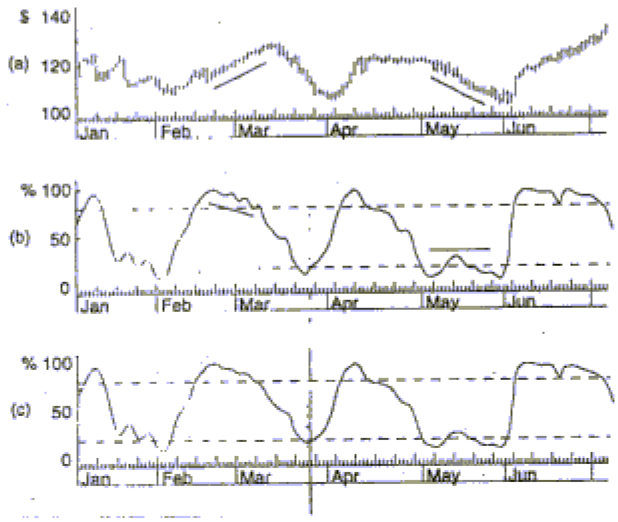
Stochastics

- *Stochastic Oscillator* is a signal designed to anticipate sudden reversals in market value (George C. Lane).
- A well-known sock market phenomenon:
  A market top, or high point for a particular stock, is usually indicated by daily closing prices that tend to cluster around the high value for the stock.

  A market bottom is indicated when daily closing prices cluster around the low value of the stock.
- *Stock prices tend to reverse their trends during a top (or bottom) period*.
- We can anticipate reversals by detecting when a stock is at (or near) its limit.
- We can develop such an indicator by comparing the current closing price of a stock, with its highest high and lowest low values over a period of time.
- *Lane's indicators* are a mathematical comparison over some fixed period of time (5 - 14 days) of the closing value...(illegible -cut off in photocopy)

%K - Comparison for 14 days
%D - Three-Day smoothed version of %K indicator.

- Daily closing prices of a stock:

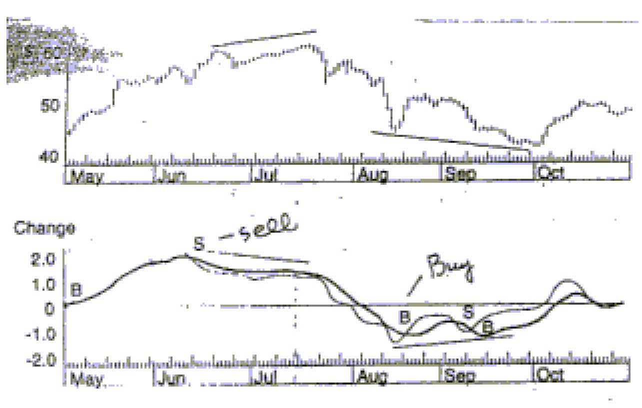


(a) Sell periods indicated by stochastic.
(b) Buy period indicated at right. Indicators have gone below 20%.
(c) A slow stochastic indicator tends to smooth data (uses three-day moving average of the %D indicator.)

- A stock is considered overbought when the stochastic indicator goes above 80%.
- A stock is considered oversold when the indicator goes below 20%.

Moving Average Convergence/Divergence

- *Moving Average Convergence/Divergence* or MACD measures the trend of a stock over a period of time.



Top graph: Closing price of a stock

Bottom graph: MACD computed.

- Note in the graph that:
  - Buy signals tend to precede periods of increasing stock value.
  - Sell signals tend to precede periods of decreasing stock value.

- To be successful in the capital management business, one must beat the market, i.e., to forecast what the market will do in the near future.

- Is the market *chaotic*?... and hence unpredictable?

- Yet models employing mathematical indicators have met with success.

- Barr, Loick of LBS Capital Management, and Fishman of Eckard College showed success at predicting the Standard & Poors (S&P) 500 Index five days into the future.

Network Architecture

- Backpropagation

- n inputs - where n = number of market indicators used.

- Normalized indicator values prior to training (BPN requires input values $\in [0,1]$.)

- Number of layers - 3 ? Indicate success with 4.

- Output - a single linear unit - a scaled prediction of the amount of change in the S&P 500 Average, five days into the future.

- Exercise 5.1: In their articles, Barr and Loick do not provide any specific guidance with regard to selecting indicators (beyond the four described here) that might improve the performance of the network, although they do indicate that their best network contained a total of 26 indicators. From your understanding of the BPN, describe the selection criteria you would apply to determine if a financial indicator could improve the performance of the network.

The Network

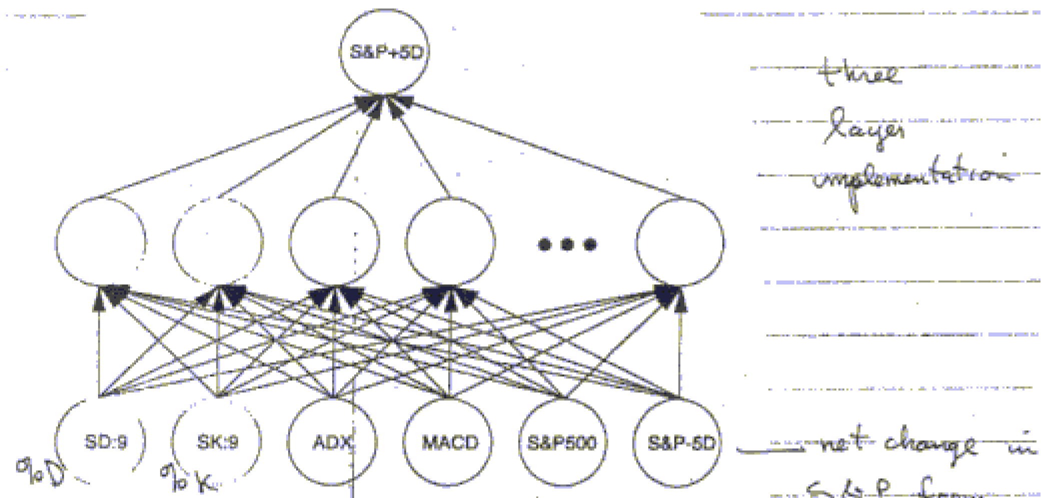


Figure 5.6 The three-layer BPN architecture used at LBS Capital Management to predict the S&P 500 average five days in the future is depicted. In this example network, the indicators used are the slow %D and slow %K stochastic indicators, using $n = 9$, an 18-day ADX, and histogram (difference) MACD values. Also used as input to the network are the current value of the S&P 500, as well as the net change in the S&P 500 value from five days prior. The output of the network is a continuously variable signal that is interpreted as a scaled estimate of the change expected in the market five days hence. *Source: Adapted from* Using neural networks in market analysis [4]. *Used with permission. Copyright ©1991, Technical Analysis, Inc.*

- Primary source of data - recent history of S&P 500.
- Training examples developed from historical market data.
- Performance of the trained network - can be evaluated by comparing actual market performance with the projections made by the network using current market data.
- Initial training data to 20 exemplars (was the LBS network overwhelmed by too much data or was training on conflicting exemplars occuring?)
- Network results:

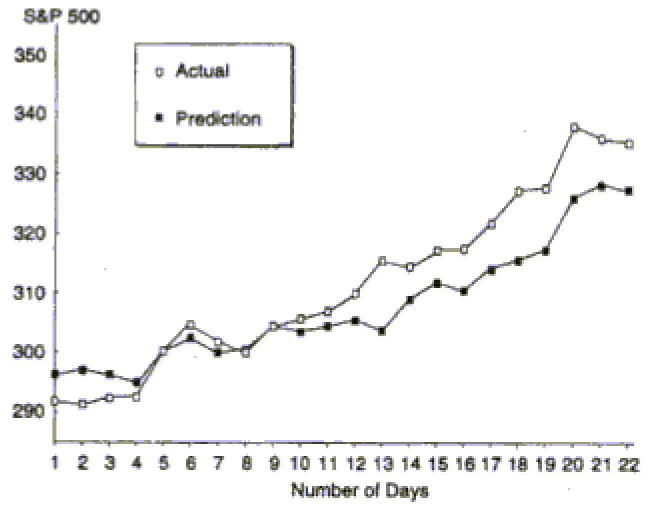


Figure 5.7 The response of the market-prediction BPN after training is shown. As this graph illustrates, the network has its lowest error in the earlier estimates, and tends to diverge from actual market values as time goes on. *Source: Adapted from* Using neural networks in market analysis [4]. *Used with permission. Copyright ©1991, Technical Analysis, Inc.*

Best estimates occur in the near future.

- Exercise 5.2: The graph in Figure 5.7 [above] shows that the neural network tends to become less accurate as the prediction date becomes more distant from the time when the training data were collected. Suggest a strategy that could be employed to reduce this error in the network's response. Explain the advantages and disadvantages of your approach.

Genetic Algorithms

- Genetic Algorithms (GA) are a parallel search method.
- A population of points encloses a local maximum of the target function after some iteration:

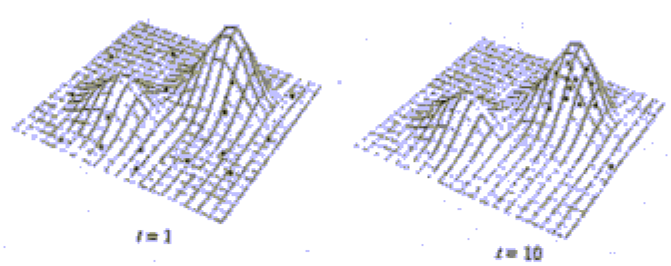


Fig. 17.1. A population of points encircles the global maximum after some generations.

- GA is a blind search - i.e., no information need be known about the search space.

- The parallel and iterative nature of the search comes at a price: Increased computation time.
- Unlike purely local search methodologies (e.g., gradient descent,) GA will not become stuck in local maxima or minima.
- Through evolution, the networking pattern of biological neural networks has been created and improved.

GA's Vs. Other Stochastic Methods

- Random Search
  Starting point $\bar{x} = (x_1, x_2, \ldots, x_n)$ vs. randomly generated and $f(x_1, x_2, \ldots, x_n)$ is computed.
- A vector $\bar{\delta} = (\delta_1, \ldots, \delta_n)$ is randomly generated and f is computed at $(x_1+\delta_1, \ldots, x_n+\delta_n)$.

  IF the value of f is thereby decreased, THEN $(x_1+\delta_1, \ldots, x_n+\delta_n)$ is taken as the *new search point* and the algorithm is started again.

  ELSE a new direction is generated.

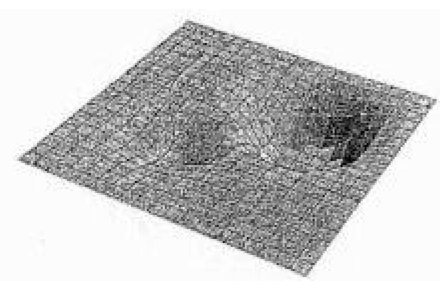


Fig. 17.2. A local minimum traps the search process

- The disadvantage of simple stochastic search is that a *local minimum* can steer the search in the wrong direction.
- One remedy is to carry out serveral independent searches.

The Metropolis Algorithm

- Annealing:
  A metal is heated until it liquifies. It is then slowly cooled until it once again changes phase, i.e., solidifies.

- Simulated Annealing:
  Variation on stochastic search.
  If a new search direction ($\delta_1, \ldots, \delta_n$) decreases the function value, it is used to update the search position.
  However, if the function value increases, this new direction is still used with probability p with,

$$p = \frac{1}{1 + e^{\frac{1}{\alpha}(f(x_1+\delta_1, \ldots, x_n+\delta_n) - f(x_1, \ldots, x_n))}}$$

  where the constant α (the "temperature") approaches zero gradually.

- At the beginning of the search process, counter-productive jumps are taken with a relatively high probability.

- In the final iterations, only productive jumps are taken.

- This methodology prevents getting trapped in a local minimum (or maximum.)

Bit-Based Descent Methods

- The problem is encoded (re-coded) so that the function f is calculated with the help of a binary string.

- Example:

  $x \longrightarrow (1 - x^2)$.

- Let $x = b_7b_6b_5\, b_4b_3b_2b_1b_0$

  where $b_7b_6b_5$ is the Whole part, and $b_4b_3b_2b_1b_0$ is the Fractional part.

  Numbers between 0 and 8 can be represented this way. A sign bit may also be added.

- A randomly chosen initial string $b_7b_6b_5\, b_4b_3b_2b_1b_0$ is generated. The function f is then computed for this point.

- A bit of the string is selected at random and flipped. For example:

  $x' = b'_7b_6b_5\, b_4b_3b_2b_1b_0$

  Here the b_7 was flipped to b'_7.

- The function f is computed at x'

  If f(x') < f(x) then x' is accepted and another iteration is performed.

- The algorithm runs until no bit flip improves the value of f.

- [Illegible ... cut off in photocopy...]

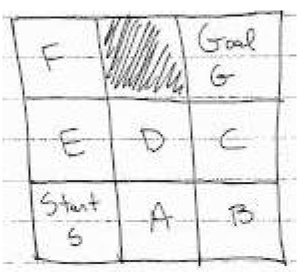
Genetic Algorithms

- Genetic algorithms are stochastic search methods managing a population of simultaneous search positions.

- GA's work with a coding of the optimization problem.

Example

- Consider the following puzzle. A robot starts at square S, and the goal square is G:



The robot can move one square at a time in any of the following directions:
North
South
East
West

As long as a move is not prevented by constraints of the board (e.g., cannot move West from D or East from F.

- We wish to reach the Goal in four moves.

- We may encode our moves as:

  Try: 00101000

Where
North = 00
South = 01

East = 10
West = 11

And hence, a series of four moves.

- GA's begin with a population of strings (search points) where population size is fixed at the beginning.

GA Operators

- *Selection*
  A string is chosen to help generate the next population based upon its fitness measure.
  For example, the string 10101111 is not likely to participate in reproduction (why not?)
- *Mutation Operator*
  A bit may be complemented with some small probability, say 0.001 during reproduction.
- *Information Exchange Operators*
  eg. crossover.

  Example: Consider two strings chosen for reproduction:
  string 1: 10100001 , (lands robot in square B)
  string 2: 00010000 , (lands robot in square F)

  A crossover point is randomly chosen, say at k = 6 (i.e., after six bits from the left)
  101000|01
  000100|00

  Then the Descendent point equals

  10100000 , sends robot to Goal.

  In the descendant string above, the first six bits are from string 1, and the last 2 are from string 2.

- *Schema*
  00 representation for all strings of length 6 with two zeros in the center positions. (e.g., 11<u>00</u>01, 00<u>00</u>10, ...)

- During the course of a GA, the best bit patterns are selected - "Schema Theorem".

Deceptive Functions

- However, some problems are difficult for GA's. These are termed *Deceptive Functions*.

This occurs when there is a correlation between optimal bits.

- e.g.,

  $(x_1, x_2, \ldots, x_n) \longrightarrow ((x_1^2, \ldots, x_n^2)/(x_1^2 + \epsilon) + \ldots + (x_1^2, \ldots, x_n^2)/(x_n^2 + \epsilon))$.

  where ϵ is a small positive constant.

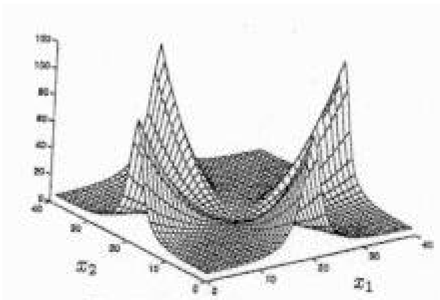


Fig. 17.4. A deceptive function

Need to approach from just the right direction.

Relatively easy problem for Gradient Descent.

- Other problems are easy. Like the so-called "*royal road*" functions.

- **Metagenetic Algorithms**
  Encode the mutation rate ("tendency to explore" regions of the search space) or the length of stochastic changes to the points in the population in the individual bit strings themselves.

- *Optimal mutation rate* sought simultaneously with the optimum of the fitness function.

Neural Nets and Genetic Algorithms

- Is it possible to use GA's to find the weights in a network?

- Is it possible to let networks evolve so that they find an optimal topology?

- There is a problem with finding the weights

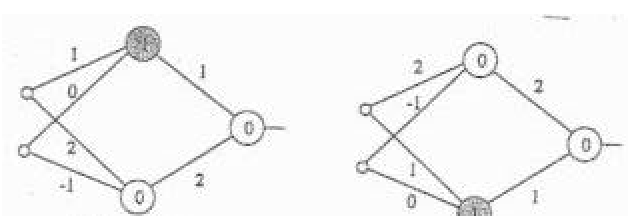


Fig. 17.6. Equivalent networks (weight permutation)

High number of symmetries for the error function...
permuting the weights...
coding for each would look very different.

Encoding/Decoding Problem with Eight Inputs

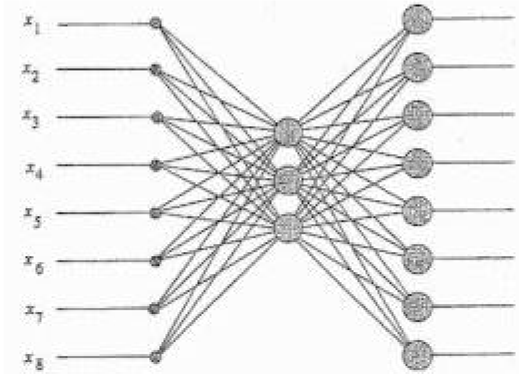


Fig. 17.8. An 8-bit encoder-decoder

- Only 1 input "on" - it must be replicated at output.

- We have,
  48 weights
  + 8 input bits
  <u>+ 3</u> bits to encode input at hidden units
  59 floating point numbers.

- Crossover through the middle of a parameter was avoided.

- Evolution of the Error curve:

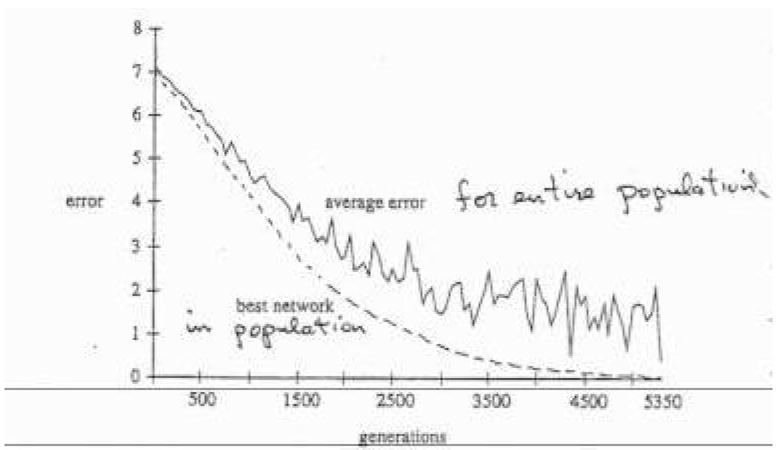


Fig. 17.9. Error function at each generation: population average and best network

After 5350 generations, error < 0.05.

- For larger networks, this is an interesting research problem.

Prisoner's Dilemma

- In this game between two players, each one decides independently whether they wish to cooperate with the other player (partner in crime) or betray him (squeal to the police.)

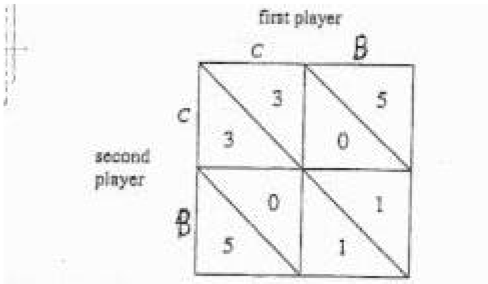


Fig. 17.10. Pay-off matrix for the prisoner's dilemma

C = Cooperate
B = Betray
Pay-off = 5 , number of years in prison.

- If the first player betrays his partner, then the first is set free, the second has 5 years in jail.
- Incentive to commit treachery:If just played once, naturally one should betray!
- More complicated if game is repeated: Iterated Prisoner's Dilemma.
- If player only stores last three games, tit-for-tat (TFT) is a successful strategy (i.e., repeat last move of your opponent.)
- Axelrod and Hamilton, 1981.