THE UNIVERSITY OF TENNESSEE KNOXVILLE

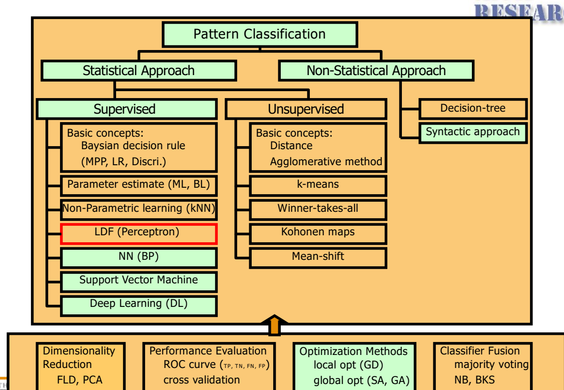
AICIP RESEARCH

ECE471-571 – Pattern Recognition

Lecture 15: NN - Perceptron

Hairong Qi, Gonzalez Family Professor
Electrical Engineering and Computer Science
University of Tennessee, Knoxville
http://www.eecs.utk.edu/faculty/qi
Email: hqi@utk.edu

AICIP RESEARCH



AICIP RESEARCH

Review - Bayes Decision Rule

$$P(\omega_j \mid x) = \frac{p(x \mid \omega_j) P(\omega_j)}{p(x)}$$

Maximum Posterior Probability: For a given x, if $P(\omega_1|x) > P(\omega_2|x)$, then x belongs to class 1, otherwise 2

Likelihood Ratio: If $\frac{p(x|\omega_1)}{p(x|\omega_2)} > \frac{P(\omega_2)}{P(\omega_1)}$, then x belongs to class 1, otherwise, 2.

Discriminant Function: The classifier will assign a feature vector x to class ω_i if $g_i(x) > g_j(x)$

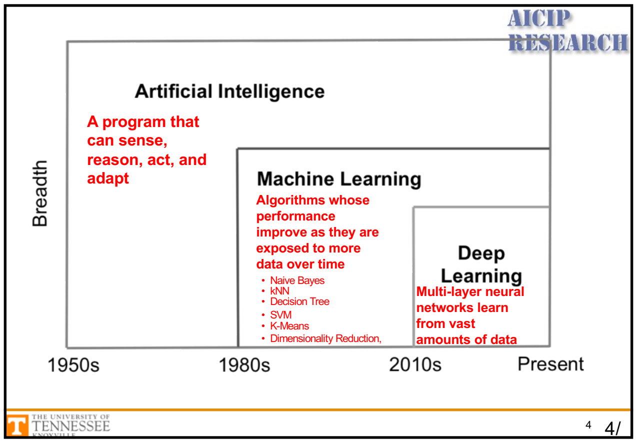
- Case 1: Minimum Euclidean Distance (Linear Machine), $\Sigma_i = \sigma^2 I$
- Case 2: Minimum Mahalanobis Distance (Linear Machine), $\Sigma_i = \Sigma$
- Case 3: Quadratic classifier , Σ_i = arbitrary

Non-parametric kNN: For a given x, if $k_1/k > k_2/k$, then x belongs to class 1, otherwise 2

Estimate Gaussian (Maximum Likelihood Estimation, MLE), Two-modal Gaussian

Dimensionality reduction

Performance evaluation ROC curve



A Bit of History

- 1956-1976
  - 1956, The Dartmouth Summer Research Project on Artificial Intelligence, organized by John McCarthy, Marvin Minsky, Nathaniel Rochester, and Claude Shannon

    > We propose that a 2 month, 10 man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College ... The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer.

  - The rise of symbolic methods, systems focused on limited domains, deductive vs. inductive systems
  - 1973, the Lighthill report by James Lighthill, "Artificial Intelligence: A General Survey" - automata, robotics, neural network
  - 1976, the AI Winter
- 1976-2006
  - 1986, BP algorithm
  - ~1995, The Fifth Generation Computer
- 2006-???
  - 2006, Hinton (U. of Toronto), Bingio (U. of Montreal), LeCun (NYU)
  - 2012, ImageNet by Fei-Fei Li (2010-2017) and AlexNet

https://en.wikipedia.org/wiki/Dartmouth_workshop
https://en.wikipedia.org/wiki/Lighthill_report

THE UNIVERSITY OF TENNESSEE

A Wrong Direction

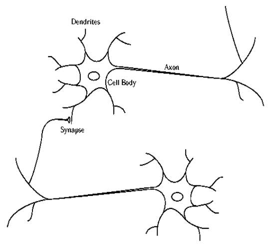
- One argument: Instead of understanding the human brain, we understand the computer. Therefore, NN dies out in 70s.
- 1980s, Japan started "the fifth generation computer research project", namely, "knowledge information processing computer system". The project aims to improve logical reasoning to reach the speed of numerical calculation. This project proved an abortion, but it brought another climax to AI research and NN research.

THE UNIVERSITY OF TENNESSEE 6

Biological Neuron

AICIP RESEARCH

- Dendrites: tiny fibers which carry signals to the neuron cell body
- Cell body: serves to integrate the inputs from the dendrites
- Axon: one cell has a single output which is axon. Axons may be very long (over a foot)
- Synaptic junction: an axon impinges on a dendrite which causes input/output signal transitions



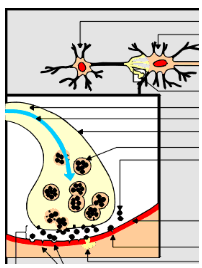
THE UNIVERSITY OF TENNESSEE 7

Synapse

http://faculty.washington.edu/chudler/chnt1.html

AICIP RESEARCH

- Communication of information between neurons is accomplished by movement of chemicals across the synapse.
- The chemicals are called neurotransmitters (generated from cell body)
- The neurotransmitters are released from one neuron (the presynaptic nerve terminal), then cross the synapse and are accepted by the next neuron at a specialized site (the postsynaptic receptor).

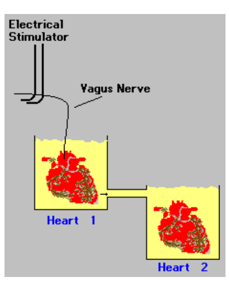


THE UNIVERSITY OF TENNESSEE 8

The Discovery of Neurotransmitters

AICIP RESEARCH

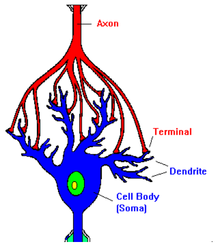
- Otto Loewi's Experiment (1920)
- Heart 1 is connected to vagus nerve, and is put in a chamber filled with saline
- Electrical stimulation of vagus nerve causes heart 1 to slow down. Then after a delay, heart 2 slows down too.
- Acetylcholine



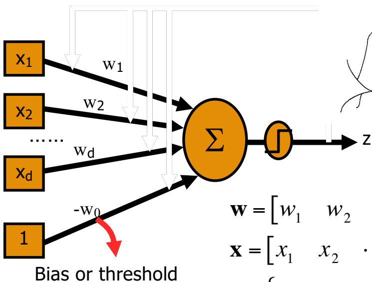
THE UNIVERSITY OF TENNESSEE 9

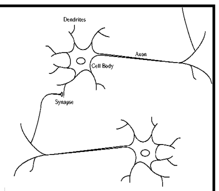
Action Potential

- When a neurotransmitter binds to a receptor on the postsynaptic side of the synapse, it results in a change of the postsynaptic cell's excitability: it makes the postsynaptic cell either more or less likely to fire an action potential. If the number of excitatory postsynaptic events are large enough, they will add to cause an action potential in the postsynaptic cell and a continuation of the "message."
- Many psychoactive drugs and neurotoxins can change the properties of neurotransmitter release, neurotransmitter reuptake and the availability of receptor binding sites.



Perceptron





$$\mathbf{w} = \begin{bmatrix} w_1 & w_2 & \cdots & w_d \end{bmatrix}$$

$$\mathbf{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_d \end{bmatrix}$$

$$z = \begin{cases} 1 & \text{if } w^T x - w_0 > 0 \\ 0 & \text{otherwise} \end{cases}$$

Perceptron

- A program that learns "concepts" based on examples and correct answers
- It can only respond with "true" or "false"
- Single layer neural network
- By training, the weight and bias of the network will be changed to be able to classify the training set with 100% accuracy

Perceptron Criterion Function

AICIP RESEARCH

$$J_p(\mathbf{a}) = \sum_{\mathbf{y} \in Y} (-\mathbf{a}^T \mathbf{y})$$ Y: the set of samples misclassified by **a**

$$\nabla J_p = \sum_{\mathbf{y} \in Y} (-\mathbf{y})$$

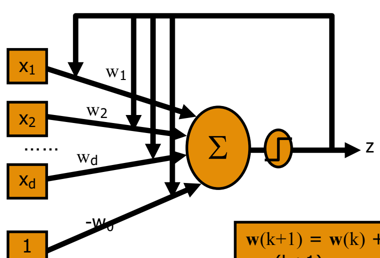
$$\mathbf{a}(k+1) = \mathbf{a}(k) + \eta(k) \sum_{\mathbf{y} \in Y_k} \mathbf{y}$$ Gradient descent learning

$\mathbf{a} = [w_1\ w_2\ \ldots\ w_d\ -w_0]$ – augmentation
$\mathbf{y} = [x_1\ x_2\ \ldots\ x_d\ \ 1]$

THE UNIVERSITY OF TENNESSEE KNOXVILLE 13

Perceptron Learning Rule

AICIP RESEARCH



$\mathbf{w}(k+1) = \mathbf{w}(k) + \Sigma(T - z)\ \mathbf{x}$
$-w_0(k+1) = -w_0(k) + \Sigma(T - z)$

$$\mathbf{a}(k+1) = \mathbf{a}(k) + \eta(k) \sum_{\mathbf{y} \in Y_k} \mathbf{y}$$

T is the expected output
z is the real output

THE UNIVERSITY OF TENNESSEE KNOXVILLE 14

Training

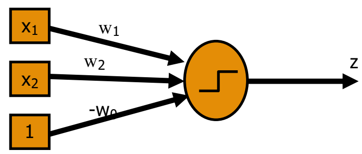
AICIP RESEARCH

- Step1: Samples are presented to the network
- Step2: If the output is correct, no change is made; Otherwise, the weight and biases will be updated based on perceptron learning rule
- Step3: An entire pass through all the training set is called an "epoch". If no change has been made for the epoch, stop. Otherwise, go back Step1

THE UNIVERSITY OF TENNESSEE KNOXVILLE 15

AICIP RESEARCH

Exercise (AND Logic)



x1	x2	T
0	0	0
1	0	0
0	1	0
1	1	1

w1 w2 $-w_0$

THE UNIVERSITY OF TENNESSEE

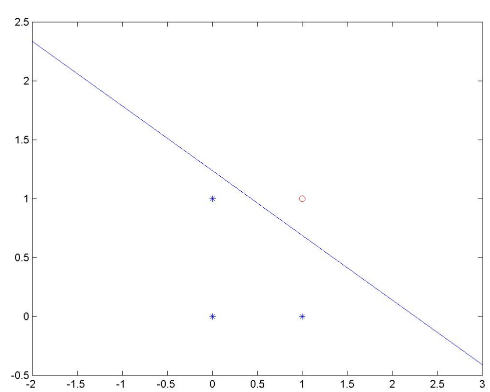
16

```
% demonstration on perceptron
% AND gate
%
% Hairong Qi
input = 2;
tr = 4;

w = rand(1,input+1);      % generate the initial weight vector
x = [0 0; 1 0; 0 1; 1 1]; % the training set (or the inputs)
T = [0; 0; 0; 1];         % the expected output

% learning process (training process)
finish = 0;
while ~finish
   disp(w)
   for i=1:tr
      z(i) = w(1:input) * x(i,:)' > w(input+1);
      w(1:input) = w(1:input) + (T(i)-z(i))* x(i,:);
      w(input+1) = w(input+1) - (T(i)-z(i));
   end
   if sum(abs(z-T')) == 0
      finish = 1;
   end
   disp(z)
   pause
end
disp(w)
```

17



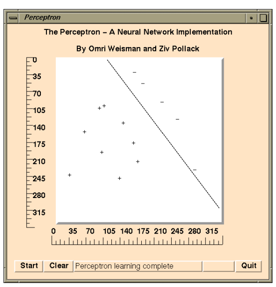
Limitations

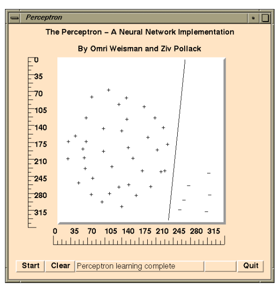
AICIP RESEARCH

- The output only has two values (1 or 0)
- Can only classify samples which are linearly separable (straight line or straight plane)
- Can’t train network functions like XOR

Examples

AICIP RESEARCH





Examples

AICIP RESEARCH