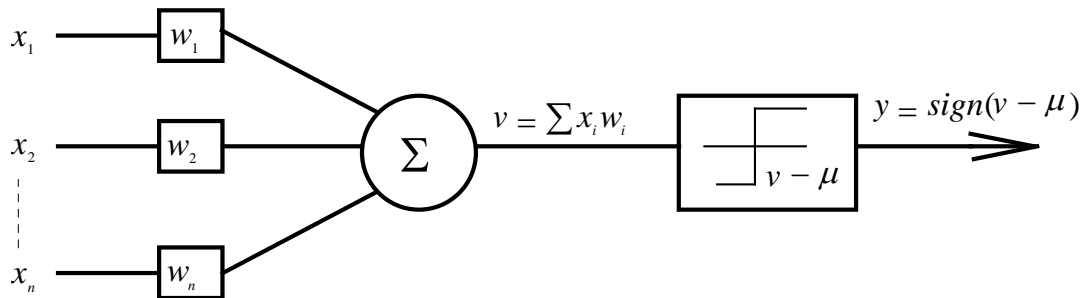


- 1.3. Alguns aspectos históricos
- 1.4. Principais Conceitos
 - 1.4.1. Definições
 - 1.4.2. Alguns tipos de Arquitetura

Revisão da Aula-1



“*Neurônio de McCulloch-Pitts.*” (1943)

Considerando uma unidade j em uma dada rede, a resposta dessa unidade é expressa por:

$$y_j = \text{sgn}(v_j - \mu_j) \quad (1)$$

onde a função *sinal* (sgn) assume dois valores:

$$\text{sgn}(x) = \begin{cases} 1 \rightarrow \text{se } x \geq 0 \\ -1 \rightarrow \text{se } x < 0 \end{cases}$$

μ é o limiar (“bias”) e,

$$v_j = \sum_{i=1}^n w_{j,i} x_{j,i} \quad (2)$$

1.3 - Aspectos Históricos

Na época do Neurônio de McCulloch-Pitts não existiam computadores para simulação, as implementações eram físicas (*hardware*), mas os componentes não eram adequados e robustos o suficiente para uma boa implementação.

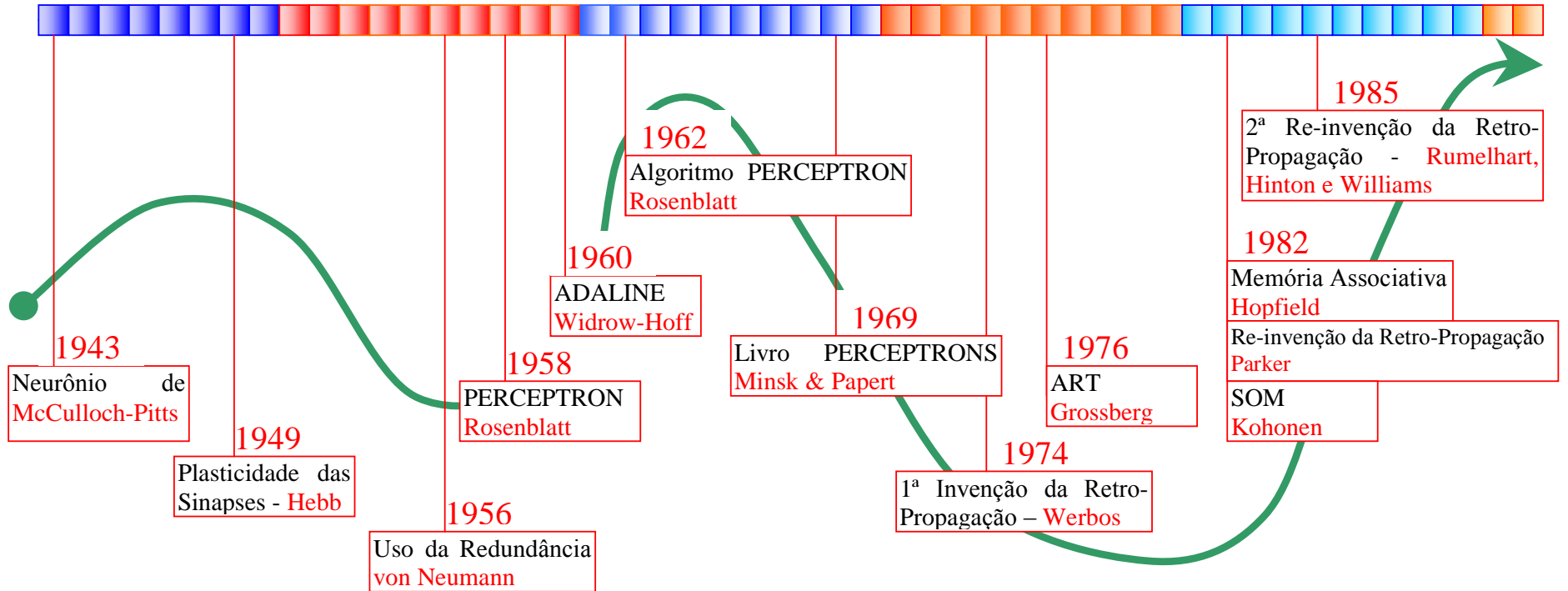
Essas dificuldades, que persistiram até o final da década de 50, diminuíram o entusiasmo inicial, gerando o que ficou marcado como o primeiro período de ceticismo, que durou até 1958, quando Rosenblatt apresentou o seu projeto de um “*computador neural*” denominado PERCEPTRON.

É importante destacar que a solução do problema de se construir máquinas confiáveis com componentes não confiáveis começou a surgir na metade da década de cinqüenta, quando John von Neumann (1956) introduziu o conceito de *redundância*, originando a geração dos atuais computadores digitais, que vieram a “*vencer a corrida*” disputada com a linha de “*cérebros eletrônicos*”.

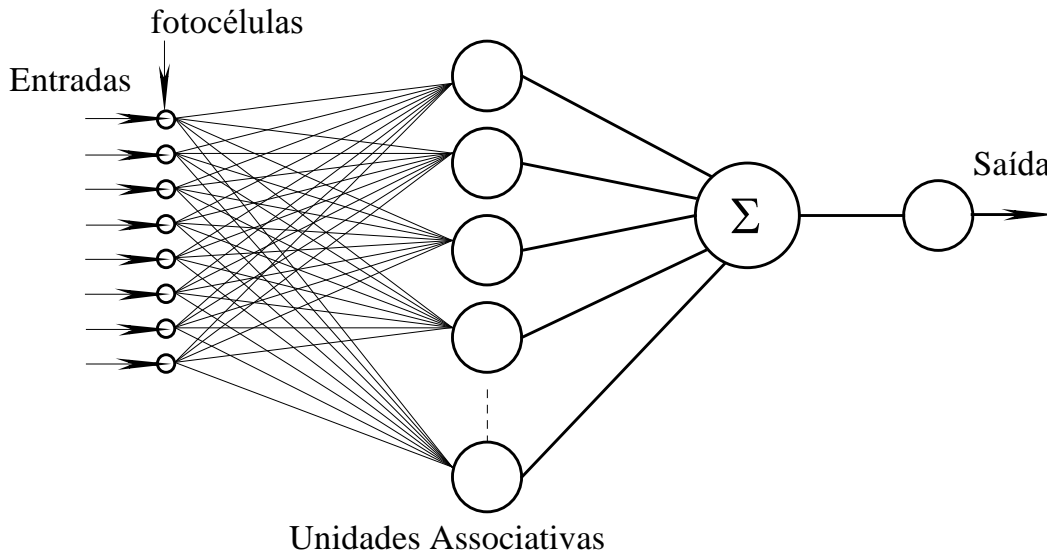
O PERCEPTRON (termo que sobrevive até hoje), que surgiu nesse período, era uma rede, de uma única camada, contendo unidades associativas alimentadas por 400 células fotoelétricas. Entre 1960 e 1962 o grupo de Rosenblatt se concentrava no problema do treinamento do PERCEPTRON, tendo conseguido provar a convergência de um algoritmo de aprendizado.

Aproximadamente na mesma época, Widrow e Hoff (1960), que seguiam por um caminho diferente do de Rosenblatt, introduziram o conceito que foi batizado como: ADALINE (sigla de *AD*Aptive *LI*Near *E*lement, posteriormente chamado por alguns autores como *AD*Aptive *LI*near *NE*uron).

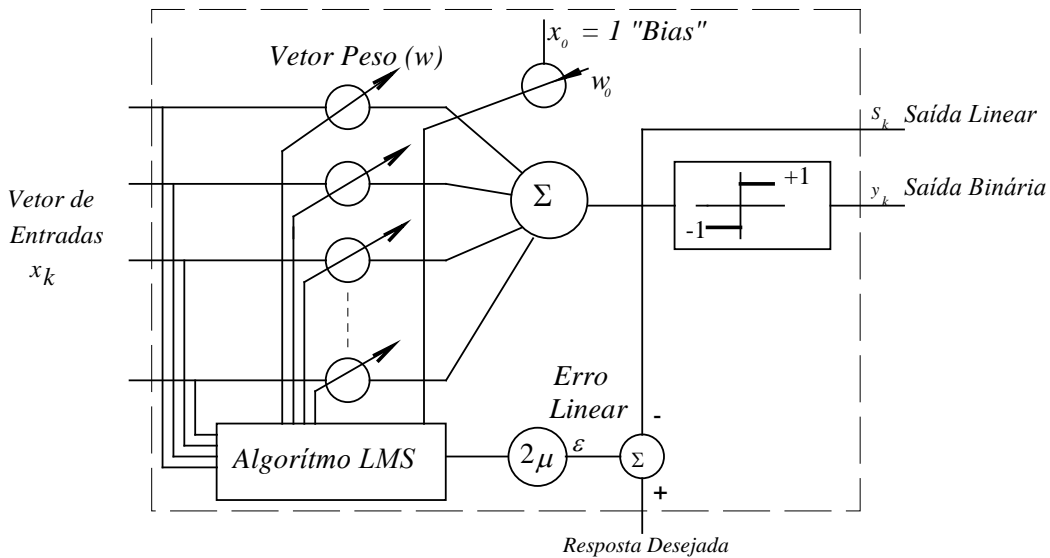
A composição de uma rede de múltiplos ADALINES's origina o chamado “MADALINE”.



O PERCEPTRON original era constituído por 400 células fotoelétricas recebendo estímulos óticos primários e interligadas a elementos processadores que, dependendo da entrada, produziam uma determinada resposta.



PERCEPTRON de Rosenblatt.

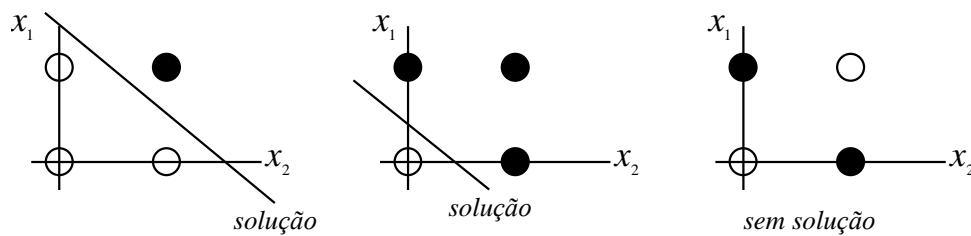


ADALINE de WIDROW (1960).

Do mesmo modo que o PERCEPTRON, o ADALINE possuía apenas uma única camada de unidades. A diferença é que o erro era baseado na saída linear. Apesar dos bons resultados com os primeiros experimentos, constatou-se que uma única camada não é capaz de implementar qualquer função, sendo limitada a solução de funções *linearmente separáveis*.

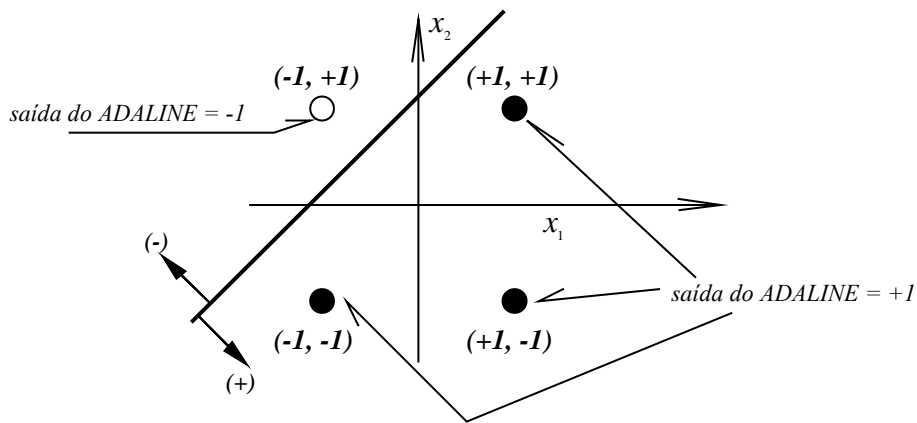
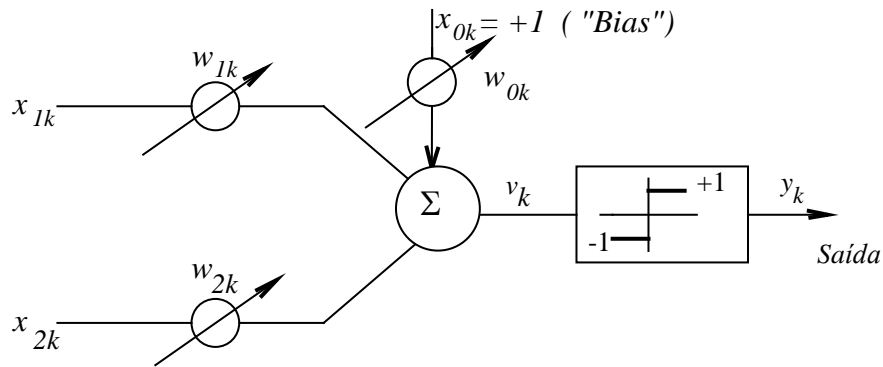
Durante muito tempo a dificuldade dessa limitação, abordada no livro PERCEPTRONS, que foi publicado em 1969 por Minsky e Papert, se tornou o maior empecilho para o desenvolvimento do campo de RNA. Nesse trabalho eles demonstraram que o teorema de Rosenblatt se aplicava somente a problemas que a rede era capaz de resolver, não sendo capaz de implementar mesmo diversas funções simples.

A Fig. a seguir mostra algumas funções para um sistema com duas entradas (x_1 e x_2), demonstrando que não é possível, para esse sistema, representar a função XOR (o “ou-exclusivo”).



Discriminador Linear Não Implementa Função "XOR".

A próxima figura mostra um ADALINE com duas entradas e a representação de sua resposta na forma de uma *linha de separação* no espaço de padrões.



Linha de Separação de um ADALINE de duas entradas.

A saída da unidade acima é ativada quando $v_k \geq 0$. No caso da unidade com duas entradas (x_1 e x_2), e no limite quando $v_k = 0$ se tem:

$$v = x_1 w_1 + x_2 w_2 + w_0 = 0$$

$$\text{portanto} \rightarrow x_2 = -\frac{w_1}{w_2} x_1 - \frac{w_0}{w_2}$$

$$\text{onde} \rightarrow \text{declividade} = -\frac{w_1}{w_2} \quad \text{e,}$$

$$\text{Interceptação} = -\frac{w_0}{w_2}$$

O ADALINE e o PERCEPTRON foram responsáveis por um grande ciclo de entusiasmo que, no entanto, esbarrando na dificuldade da *separação linear*, quase desapareceu quando foram praticamente cortados todos os recursos +destinados às pesquisas em redes neurais.

Muito tempo depois se constatou que grupos de elementos simples organizados em camadas, semelhantes aos PERCEPTRONS mas utilizando um novo algoritmo de aprendizado, a *regra delta* de aprendizado, podiam implementar todas as funções básicas dos computadores digitais (na época se pensava ainda na utilização de *funções lógicas* para o desenvolvimento do *cérebro eletrônico*).

Apesar do arrefecimento provocado pelo corte dos investimentos em redes neurais e, portanto com muita dificuldade, alguns pesquisadores prosseguiram seu trabalho. Somente para ilustrar destacam-se os trabalhos de:

- James Anderson (1968), que trabalhava em um modelo baseado em modelos biológicos da memória e de reconhecimento, nos quais as forças de ligação sinápticas são fortalecidas quanto mais freqüentemente são ativadas.
- Tuevo Kohonen (1974-1982), que desenvolveu o conceito de uma rede chamada “*mapa auto-organizável*”. A contribuição de Kohonen define o conceito de “aprendizado competitivo”, no qual as unidades competem para responder a determinada entrada e o elemento vencedor tem os pesos de suas entradas modificados para responder a valores próximos dessa entrada, cada vez com maior força.
- Stephen Grossberg (1967-1969), cujos estudos procuravam utilizar dados neurológicos na construção de novos modelos em computação neural. Seus trabalhos se focalizaram no estudo dos reflexos condicionados, baseando-se nos postulados de Hebb (1949) sobre o aprendizado.

- Werbos (1974), desenvolveu e apresentou na sua tese de doutorado um algoritmo de aprendizado que é o método atualmente denominado de retro-propagação (ou “propagação para trás”). Infelizmente esse algoritmo não chegou a ser disseminado na época.
- Grossberg (1976), desenvolveu a Teoria da Ressonância Adaptativa (Adaptive Resonance Theory – ART) que trata de implementações neurais com algoritmos auto-organizáveis de agrupamento de padrões.
- Parker (1982), também ainda sem uma merecida divulgação, reinventou esse mesmo algoritmo.
- John Hopfield (1982), apresentou um trabalho que é uma particularização do trabalho de Grossberg e que ficou conhecido como “redes associativas em linha transversal”, constituídas por elementos interligados buscando o aprendizado com “um mínimo de energia”. O trabalho de Hopfield originou um novo ciclo de entusiasmo e progresso nessa outra linha.
- Rumelhart, Hinton e Williams (1985) mais uma vez reinventaram o algoritmo de retro-propagação e, somente a partir daí, começaram a se desenvolver muitas pesquisas dentro dessa linha. Atualmente muitos grupos em praticamente todas as universidades e centros de pesquisa desenvolvem trabalhos no campo de redes neurais artificiais, ainda baseados nos estudos acima.

1.4 - Principais Conceitos

1.4.1. Definições

Ação	: Processo durante o qual uma rede dá resposta a um conjunto de parâmetros de entrada...
Arquitetura	: É o modo de organização das unidades da rede, em conjunto com a função de transferência, fluxo de sinais e ainda, às vezes, em conjunto com o algoritmo de treinamento que a distingue de outras.
Limiar (Bias)	: Valor que corresponde a “limiar de disparo” e promove uma “transformação afim...”
Camada	: Conjunto de unidades em paralelo que recebem sinais sincronizadamente...
Conexão	: Representa uma sinapse...
Função de ativação, ganho, ou transferência	: Função resposta da unidade (<i>função sinal, sigmóide, tangente hiperbólica, etc.</i>)
Peso	: Multiplicador das entradas em uma unidade que representa uma “resistência/ganho sináptico”
Tabela de Treinamento	: Conjunto de dados entrada/saída para treinamento de uma rede
Treinamento	: Processo de apresentação de exemplos e alteração de pesos para aprendizado
Unidade	: Representação artificial de um neurônio

PADRÃO

“RNA’s servem para *classificar padrões!*”

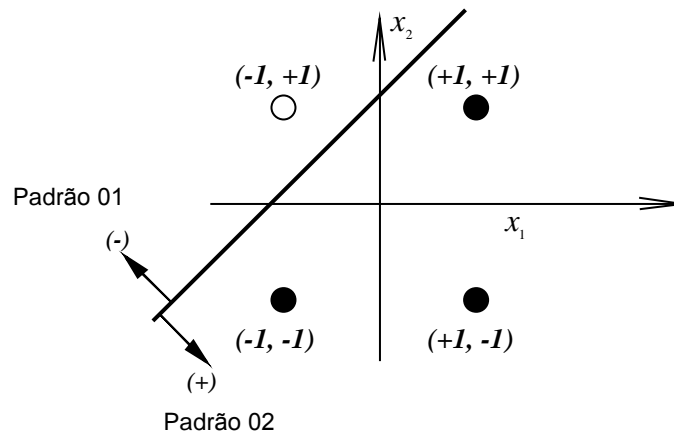
De fato, o que elas fazem é “*mapear*” um conjunto de dados, construindo uma representação interna!

O que é um “*padrão*” no campo de RNA?

Um padrão pode ser um *estado*, uma *condição*, uma *característica*, etc. de um determinado processo, objeto, população, etc.

Por exemplo:

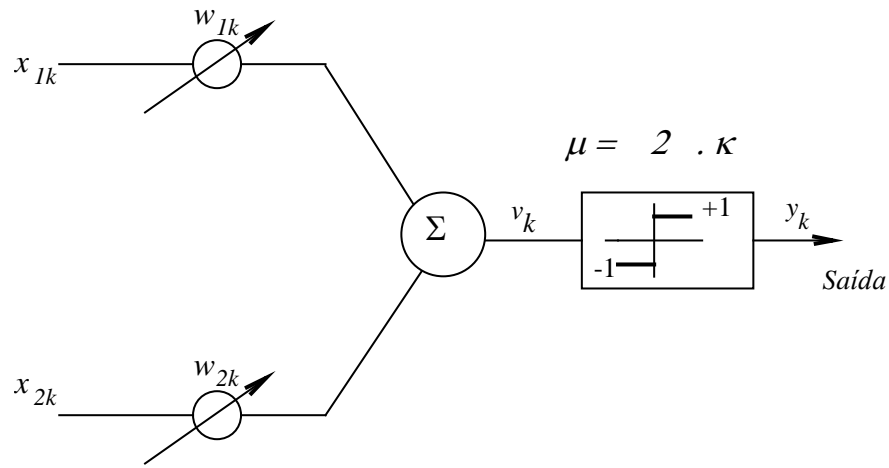
- A pressão (p) e a temperatura (T), que definem um estado de um determinado sistema, definem *um padrão para propriedades termodinâmicas* → *padrão 01* (p_1, T_1)
- A idade (i), o sexo (s) e a altura (h), características de uma dada população, podem representar padrões a serem classificados → *padrão p* (i_p, s_p, h_p)



Segundo Exemplo

Implementação de funções “E”, “OU” e outras com uma unidade de 2 entradas

Considere a Unidade abaixo:



Para resolver esse problema vamos utilizar a planilha Excel: *Separação Linear-Aula-2.xls*

A unidade pode ser treinada para tentar reproduzir qualquer uma das funções a seguir (uma de cada vez). Vamos verificar...

OU

x1	x2	y
0	0	0
0	1	1
1	0	1
1	1	1

x2

x1	x2	y
0	0	0
0	1	1
1	0	0
1	1	1

E

x1	x2	y
0	0	0
0	1	0
1	0	0
1	1	1

x1

x1	x2	y
0	0	0
0	1	0
1	0	1
1	1	1

Será que ela é capaz de “aprender” a função “Ou-Exclusivo”?

OU-Exclusivo

x1	x2	y
0	0	0
0	1	1
1	0	1
1	1	0

1.4.2. Algumas “Arquiteturas”

Quase pode ser afirmado que *“cada aplicação, ou melhor, cada tipo de problema, pode requerer uma arquitetura específica para sua melhor solução”*. Por esse motivo podem ser encontradas inúmeras arquiteturas de RNA, muitas definidas apenas por uma classificação do fluxo de sinais, da organização das unidades, da forma da função de transferência, por alguma característica específica do algoritmo de treinamento, ou ainda por outros motivos. O processo de construção de uma rede ainda é uma atividade repleta de “heurísticas”.

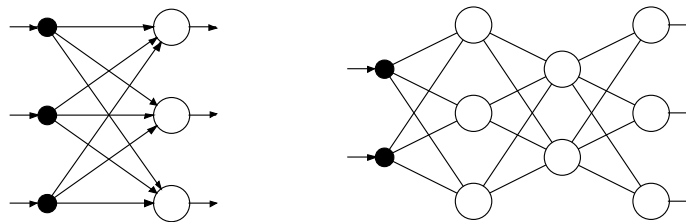
De um modo geral as redes neurais artificiais formam arranjos regulares de unidades que representam os neurônios. Essas unidades, uniformemente organizadas, são associadas em paralelo e em camadas, como se constituíssem um cérebro em um estágio primitivo de desenvolvimento, ou seja, antes de adquirir *“significado”*. As *forças de ligação* entre essas unidades, denominadas de *“pesos”*, após serem *“criadas”* tendo valores pequenos e escolhidos aleatoriamente, são ajustadas num processo de *“aprendizado”*.

Essas unidades se comunicam e a transmissão dos sinais entre elas pode se resumir em: amplificação ou atenuação linear dos sinais de entrada através da sua multiplicação pelos pesos; soma desses sinais ponderados; e, emissão de um sinal de resposta filtrado por meio da função de transferência (em geral representada por uma tangente hiperbólica ou uma sigmóide).

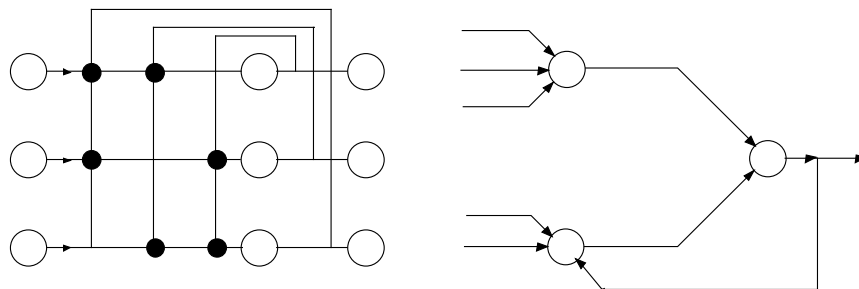
Com base nesses conceitos é que se define a ***arquitetura da rede***: que é *o modo de organização de suas unidades em conjunto com o fluxo dos sinais e ainda às vezes em conjunto com o processo de aprendizado envolvido*.

Para cada tipo de arquitetura se encontram muitos exemplos de aplicações, como por exemplo, o *reconhecimento de padrões* por meio de PERCEPTRONS, a *previsão do tempo* utilizando MADALINE, a *codificação e compressão de imagens*, etc., todas essas representando a utilização de redes

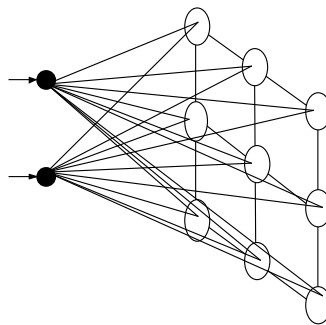
alimentadas adiante (“feed-forward”); o *posicionamento de um braço robótico* utilizando os mapas auto-organizáveis (rede competitiva) de Kohonen; e o conceito de *memória associativa* utilizando redes em realimentação (“feedback”) de Hopfield, etc. Muitos outros conceitos, que até poderiam ser encaixados nos tipos de arquitetura descritos anteriormente, são tratados como de arquitetura diferente por introduzirem aperfeiçoamentos significativos; nesse caso são batizados conforme sua aplicação ou pelo nome de seu idealizador. A Fig. seguinte ilustra os três tipos de arquitetura mais comuns: *redes em “feed-forward”*; *redes em “feedback”*; e *mapas auto-organizáveis*.



(a) Redes Alimentadas Adiante (“feed-forward”)



(b) Redes com realimentação (“feedback”)



(c) “Mapas Auto-Organizáveis” (Rede competitiva)

Tipos de Arquitetura de Redes.

Questões:

1. A qual evento se atribui o marco inicial do campo de *RNA*?
2. O que era o *PERCEPTRON* de Rosenblatt?
3. Descreva, com suas palavras, o comportamento que sempre caracterizou o desenvolvimento do campo de *RNA*?
4. Represente o tipo de separação de padrões que um “neurônio de McCulloch” com duas entradas pode fazer. Represente pelo menos duas situações.
5. Qual a maior diferença entre o *PERCEPTRON* e o *ADALINE*.
6. O que você entende por *Limiar*?
7. Defina o que é *padrão* no campo de *RNA*.
8. Relacione três (3) tipos de redes diferentes pelo fluxo de sinais.
9. Utilizando a Planilha Excel fornecida, teste todas as possíveis combinações de entradas para a unidade singular e relate o observado.