6
Clusterização

Segundo JAIN et al., 1999, clusterização é a classificação não-supervisionada de dados, formando agrupamentos ou clusters. A análise de clusters envolve, portanto, a organização de um conjunto de padrões em clusters, de acordo com alguma medida de similaridade. Usualmente representados na forma de vetores de atributos ou pontos em um espaço multidimensional, espaço de atributos

6.1
Mapas auto-organizáveis

O objetivo de um algoritmo para aprendizagem auto organizada ou aprendizagem não-supervisionada é descobrir padrões significativos ou características nos dados de entrada e realizar de tal maneira que não necessite de ajuda de um "professor", modifica-se repetidamente os pesos sinápticos de uma rede neural em reposta a padrões de ativação e de acordo com regras preestabelecidas, até que se desenvolva uma configuração final. Para fazer isto, o algoritmo dispõe de um conjunto de regras de natureza local, que o capacita a aprender a calcular um mapeamento de entrada-saída com propriedades desejadas especificas. O termo local significa que a modificação aplicada ao peso sináptico de um neurônio é confinada à vizinhança imediata daquele neurônio. A modelagem das estruturas de rede usadas para aprendizagem auto-organizada tende a seguir as estruturas neurobiológicas de uma maneira muito mais extensa do que na aprendizagem supervisionada (HAYKIN, 2001).

A organização da rede acontece em dois níveis diferentes que interagem entre si na forma de um laço de realimentação. Os dois níveis são:

- Atividade. Certos padrões de atividade são produzidos por uma determinada rede em resposta a sinais de entrada

- Conectividade. Forças de conexão (pesos sinápticos) da rede são modificadas em resposta a sinais neurais dos padrões de atividade, devido a plasticidade sináptica.

A realimentação entre as modificações nos pesos sinápticos e as modificações nos padrões de atividade deve ser positiva para se obter auto-organização (em vez de estabilização) da rede. Segundo Von der Malsburg 1990a existem alguns princípios que regem a auto-organização:

- Modificações dos pesos sinápticos tendem a se auto-amplificar. Este processo é restrito pela exigência que as modificações dos pesos sinápticos devam ser baseadas em sinais disponíveis localmente.

- A limitação dos recursos leva à competição entre sinapses e com isso à seleção das sinapses que crescem mais vigorosamente às custas das outras. Estes princípio também é possibilitado pela plasticidade sináptica. Para se estabilizar o sistema deve haver alguma forma de competição por recursos limitados, por exemplo, número de entradas e recursos de energia. Especificamente, um aumento na forca de algumas sinapses da rede deve ser compensado por uma redução em outras sinapses. Consequentemente, apenas sinapses "bem-sucedidas" podem aumentar, enquanto que as não tão bem-sucedidas tendem a se enfraquecer e eventualmente desaparecer.

- As modificações em pesos sinápticos tendem a cooperar. Uma única sinapse por si só não pode produzir eficientemente eventos favoráveis. Para fazer isso é necessária a cooperação entre um conjunto de sinapses que convirjam para um neurônio particular e que carreguem sinais coincidentes suficientemente fortes para ativar aquele neurônio. A presença de uma sinapse vigorosa pode reforçar o ajuste de outras sinapses apesar da competição global da rede.

- Ordem e estrutura nos padrões de informação representam informação redundante que é adquirida pela rede neural artificial na forma de conhecimento, que é um pré-requisito necessário para a aprendizagem auto-organizada (BARLOW, 1989). Parte deste conhecimento pode ser obtido por observações dos parâmetros estatísticos como a média, a variância e a matriz de correlação dos dados de entrada.

Estas regras são baseadas na aprendizagem competitiva e os neurônios da saída competem entre si para serem ativados ou disparados com o resultado que apenas um neurônio de saída, ou um neurônio por grupo, está ligado em um instante de tempo. Um neurônio de saída que vence a competição é chamado de um neurônio vencedor leva tudo ou simplesmente neurônio vencedor.

Em um mapa auto-organizável, os neurônios estão colocados em nós de uma grade que é normalmente uni ou bi-dimensional. Mapas de dimensionalidade mais alta são também possíveis mas não tão comuns. Os neurônios se tornam seletivamente sintonizados a vários padrões de entrada (estímulos) ou classes de padrões de entrada no decorrer de um processo de aprendizagem. As localizações dos neurônios assim sintonizados se tornam ordenadas entre si de forma que um sistema de coordenadas significativo para diferentes características de entrada é criado sobre a grade (KOHONEN, 1990a). Um mapa auto-organizável é, portanto, caracterizado pela formação de um mapa topográfico dos padrões de entrada no qual as localizações espaciais (i.e. coordenadas) dos neurônios na grade são indicativas das características estatísticas intrínsecas contidas nos padrões de entrada, daí o nome “mapa auto-organizável”.

Devido a um mapa auto-organizável ser inerentemente não linear, ele pode ser visto como uma generalização não linear da análise de componentes principais (RITTER, 1995)

6.1.1
Mapa Auto-Organizável de Kohonen

O modelo introduzido por Kohonen (1982) não pretende explicar detalhes neurobiológicos. O modelo captura as características essenciais dos mapas computacionais do cérebro e ainda se mantém tratável do ponto de vista

computacional. Fazem parte de um grupo de redes neurais artificiais chamado redes baseadas em modelos de competição, ou simplesmente, redes competitivas (FAUSETT, 1994).

Outra característica importante deste tipo de rede é que elas utilizam treinamento não supervisionado, onde a rede busca encontrar similaridades baseando-se apenas nos padrões de entrada. O principal objetivo dos mapas auto-organizáveis de Kohonen é agrupar os dados de entrada que são semelhantes entre si formando classes ou agrupamentos denominados *clusters.*

Em uma rede classificadora há uma unidade de entrada para cada componente do vetor de entrada. Cada unidade de saída representa um cluster, o que limita a quantidade de clusters ao número de saídas. Durante o treinamento a rede determina a unidade de saída que melhor responde ao vetor de entrada; o vetor de pesos para a unidade vencedora é ajustado de acordo com o algoritmo de treinamento.

Durante o processo de auto-organização do mapa, a unidade do cluster cujo vetor de pesos mais se aproxima do vetor dos padrões de entrada é escolhida como sendo a "vencedora". A unidade vencedora e suas unidades vizinhas têm seus pesos atualizados.

Kohonen (1990a) formulou o princípio da formação de mapas topográficos: a localização espacial de um neurônio de saída em um mapa topográfico corresponde a um domínio ou característica particular do dado retirado do espaço de entrada.

O modelo de Kohonen aparentemente é mais geral que o modelo de Willshaw-Von Der Malsburg na medida que ele é capaz de realizar compressão de dados (i.e., a redução da dimensionalidade na entrada).

Na realidade o modelo de Kohonen pertence à classe de algoritmos de codificação vetorial. O modelo produz um mapeamento topológico que localiza otimamente um número fixo de vetores (i.e., palavras de código) em um espaço de entrada de dimensionalidade mais elevada, e desse modo facilita a compressão de dados. O modelo de Kohonen pode, portanto, ser derivado de dois modos. Pode-se utilizar as ideias básicas da auto-organização, motivadas por considerações neurobiológicas, para derivar o modelo, que é a abordagem tradicional (Kohonen, 1982, 1990a, 1997a). Alternativamente, pode-se utilizar uma abordagem de quantização vetorial que usa um modelo envolvendo um codificador e um

decodificador, que é motivada por considerações da teoria de comunicação (Luttrell, 1989b, 1991a)

O principal objetivo do mapa auto-organizável (SOM : *self-organizing map*) é transformar um padrão de sinal incidente de dimensão arbitrária em um mapa discreto uni ou bidimensional e realizar esta transformação adaptativamente de uma maneira topologicamente ordenada.

A figura 6.1 mostra o diagrama esquemático de uma grade bidimensional de neurônios normalmente usados como o mapa discreto. Cada neurônio da grade está totalmente conectado com todos os nós de fonte da camada de entrada. Esta grade representa uma estrutura alimentada adiante com uma única camada computacional consistindo de neurônios arranjados em linhas e colunas. Neste caso de grade unidimensional a camada computacional consiste simplesmente de uma única coluna ou linha de neurônios.

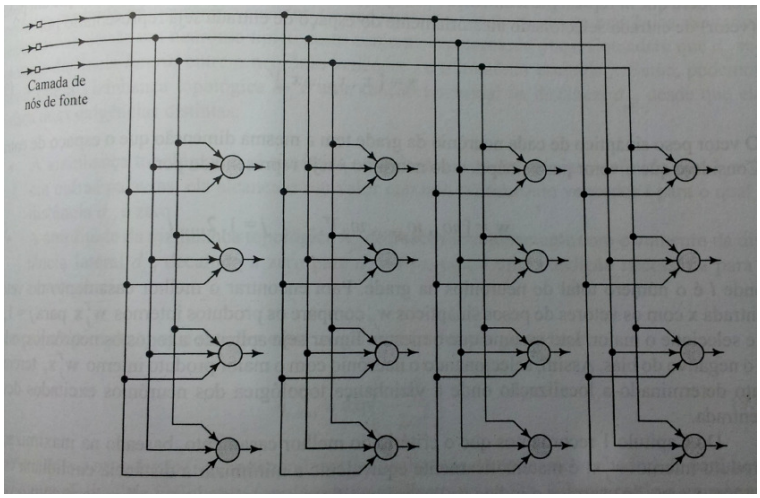

Figura 6.1.Diagrama esquemático de uma grade bidimensional de neurônios. Fonte Haykin, 2001

Cada padrão de entrada representado à grade consiste tipicamente de uma região localizada ou “foco” de atividade contra um fundo em repouso. A localização e a natureza deste foco usualmente variam de uma realização do padrão de entrada para outra. Todos os neurônios da grade devem, portanto, ser expostos a um número suficiente de diferentes realizações do padrão de entrada para assegurar que o processo de auto-organização tenha uma chance de amadurecer apropriadamente (HAYKIN, 2001)

O algoritmo responsável pela formação do mapa auto-organizável começa primeiramente inicializando os pesos sinápticos da grade. Isto pode ser feito atribuindo-lhes valores pequenos tomados de um gerador de números aleatórios, fazendo dessa forma, nenhuma organização previa é imposta ao mapa de características. Uma vez que a grade tenha sido propriamente inicializada, há três processos essenciais envolvidos na formação do mapa auto-organizável, como resumido aqui:

- Competição. Para cada padrão de entrada, os neurônios da grade calculam seus respectivos valores de uma função discriminante. Esta função discriminante fornece a base para a competição entre os neurônios. O neurônio particular com o maior valor da função discriminante é declarado vencedor da competição.
- Cooperação. O neurônio vencedor determina a localização espacial de uma vizinhança topológica de neurônios excitados, fornecendo assim a base para cooperação entre os neurônios vizinhos.
- Adaptação sináptica. Este último mecanismo permite que os neurônios excitados aumentem seus valores individuais da função discriminante em relação ao padrão de entrada através de ajustes adequados aplicados a seus pesos sinápticos. Os ajustes feitos são tais que a resposta do neurônio vencedor à aplicação subsequente de um padrão de entrada similar é melhorada.

Importante notar que a presença de redundâncias nos dados de entrada é necessário para a aprendizagem, pois fornece conhecimento.

Será apresentado uma descrição mais detalhada dos processos de competição, cooperação e adaptação sináptica.

6.1.2
O processo competitivo

Considere que *m* represente a dimensão do espaço de entrada de dados. Considere que um padrão (vetor) de entrada selecionado aleatoriamente do espaço de entrada seja representado por

$$\mathbf{x} = [\ x1\ ,\ x2\ ,\ ...\ ,\ xm\ ]^T \tag{6.1}$$

O vetor peso sináptico de cada neurônio da grade tem a mesma dimensão que o espaço de entrada. Considere o vetor peso sináptico do neurônio *j* seja representado por

$$\mathbf{w}j = [\ w1\ ,\ w2\ ,\ ...\ ,\ wjm\ ]^T \qquad j=1,\ 2,\ ...,\ l. \tag{6.2}$$

onde *l* é o número total de neurônios na grade. Para encontrar o melhor casamento do vetor de entrada **x** com os vetores de pesos sinápticos **w**j , deve-se comparar os produtos internos $\mathbf{w}^T j\ \mathbf{x}$ para *j* = 1, 2, ..., *l* e selecionar o maior. Isto assume que o mesmo limiar seja aplicado a todos os neurônios, este limiar é o negativo do bias. Assim, selecionando o neurônio com o maior produto interno $\mathbf{w}^T j\ \mathbf{x}$, tem-se de fato determinado a localização onde a vizinhança topológica dos neurônios excitados deve ser centrada.

O critério de melhor casamento baseado na maximização do produto interno $\mathbf{w}^T j\ \mathbf{x}$ é matematicamente equivalente a minimizar a distância euclidiana entre os vetores **x** e **w**j . Se for utilizado o índice i(**x**) para identificar o neurônio que melhor casa com o vetor de entrada **x**, pode-se determinar i(**x**) aplicando a condição

$$i(\mathbf{x}) = \arg\min j\ \|\mathbf{x} - \mathbf{w}j\| \qquad j = 1,\ 2,\ ...,\ l \tag{6.3}$$

que resume a essência do processo competitivo entre os neurônios. De acordo com a equação 6.3, i(**x**) é o objeto de maior atenção pois o que se busca é a identidade do neurônio *i* . O neurônio particular *i* que satisfaz esta condição é o próprio neurônio vencedor para o vetor de entrada **x**. Esta equação tem como observação:

"Um espaço continuo de entrada de padrões de ativação é mapeado para um espaço discreto de saída de neurônios por um processo de competição entre os neurônios da grade"

Dependendo da aplicação de interesse a resposta da grade pode ser tanto o índice do neurônio vencedor (i.e., sua posição na grade), como o vetor de peso sináptico que está mais próximo do vetor de entrada em um sentido euclidiano.

6.1.3
O processo cooperativo

O neurônio vencedor localiza o centro de uma vizinhança topológica de neurônios cooperativos. Um neurônio que está disparando tende a excitar mais fortemente os neurônios na sua vizinhança imediata que aqueles distantes dele, o que é intuitivamente razoável. Esta observação leva a fazer com que a vizinhança topológica em torno do neurônio vencedor *i* decaia suavemente com a distância lateral (LO et al., 1991, 1993; RITTER ET al., 1992). Sendo mais especifico considere que $h_{j,i}$ represente a vizinhança topológica centrada no neurônio vencedor *i* e que contenha um conjunto de neurônios excitados (cooperativos), sendo um neurônio típico deste conjunto representado por *j.* Considere que $d_{i,j}$ represente a distância lateral entre o neurônio vencedor *i* e o neurônio excitado *j.* Então, pode-se assumir que a vizinhança topológica $h_{j,i}$ é uma função unimodal da distância $d_{j,i}$, desde que ela satisfaça duas exigências distintas :

- A vizinhança topológica $h_{j,i}$ é simétrica em relação ao ponto máximo definido por $d_{i,j} = 0$; em outras palavras, ela alcança o seu valor Máximo no neurônio vencedor *i* para o qual a distância $d_{j,i}$ é zero.
- A amplitude da vizinhança topológica $h_{j,i}$ decresce monotonamente com o aumento da distância lateral $d_{j,i}$, decaindo a zero para $d_{j,i} \rightarrow \infty$; está é uma condição necessária para a convergência.

Uma escolha típica de $h_{j,i}$ que satisfaz estas exigências é a *função gaussiana*

$$h_{j,i(x)} = \exp (- d^2 j,i \ / 2 \sigma^2) \qquad (6.4)$$

que é invariante a translação (i.e., independentemente da localização do neurônio vencedor). O parâmetro σ é a largura efetiva da vizinhança topológica. Ele mede

o grau com o qual neurônios excitados na vizinhança do neurônio vencedor participam do processo de aprendizagem. Seu uso também faz com que o algoritmo SOM convirja mais rapidamente que com uma vizinhança topológica retangular (LO et al., 1991, 1993; RITTER ET al., 1992).

Para que a cooperação entre neurônios vizinhos se mantenha, é necessário que a vizinhança topológica $h_{j,i}$ seja dependente da distância lateral $d_{j,i}$ entre o neurônio vencedor i e o neurônio excitado j no espaço de saída em vez de ser dependente de alguma medida de distância no espaço de entrada original. No caso de uma grade unidimensional, $d_{j,i}$ é um inteiro igual a $| j - i |$. Por outro lado, no caso de uma grade bidimensional ela é definida por

$$d^2_{j,i} = \| \mathbf{r_j} - \mathbf{r}i \|^2 \tag{6.5}$$

Onde o vetor discreto **r**j define a posição do neurônio excitado j e **r**i define a posição discreta do neurônio vencedor i, sendo ambos medidos no espaço de saída discreto.

Uma outra característica única do algoritmo SOM é que o tamanho da vizinhança topológica diminui com o tempo. Esta exigência é satisfeita fazendo-se com que a largura σ da função de vizinhança topológica $h_{j,i}$ diminua com o tempo. Uma boa escolha para a dependência de σ com o tempo discreto n é o decaimento exponencial descrito por (RITTER et al., 1992; OBERMAYER et al., 1991)

$$\sigma\,(n) = \sigma 0 \exp\,(- n / \tau_1) \quad n = 0,1,2,..., \tag{6.6}$$

onde σ0 é o valor de σ na inicialização do algoritmo SOM, e τ_1 é uma constante de tempo. Consequentemente, a vizinhança topológica assume uma forma variável no tempo, como mostrado por

$$hj,i(x)\,(n) = \exp\,(- d^2j,i / 2\,\sigma^2\,(n))\ , \ n = 0,1,2,..., \tag{6.7}$$

onde σ(n) é definido pela equação 6.6. Assim, quando o tempo n (i.e., o número de iterações) aumenta a largura σ(n) decresce a uma taxa exponencial e a vizinhança topológica diminui de uma maneira correspondente. Define-se $h_{j,i(x)}\,(n)$ como a função de vizinhança.

Um outro modo útil de ver a variação da função de vizinhança $h_{j,i(x)}(n)$ em torno de um neurônio vencedor $i(\boldsymbol{x})$ é como segue (LUTTRELL, 1989a . O propósito de um $h_{j,i(x)}(n)$ largo é essencialmente correlacionar as direções das atualizações dos pesos de um grande número de neurônios excitados da grade. Quando a largura de $h_{j,i(x)}(n)$ é diminuída também diminui o número de neurônios cujas direções de atualização são correlacionadas.

6.1.4
O processo adaptativo

Neste último processo, o processo adaptativo sináptico, na formação auto-organizada de um mapa de características. Para que a grade seja auto-organizável, é necessário que o vetor de peso sináptico **w**j do neurônio *j* da grade se modifique em relação ao vetor de entrada **x**. No postulado de aprendizagem de Hebb, um peso sináptico é aumentado com uma ocorrência simultânea de atividades pré-sináptica e pós-sináptica. O uso de tal regra é muito adequada para aprendizagem associativa.

Usando o formalismo de tempo discreto, dados o vetor peso sináptico $w_j(n)$do neurônio *j* no tempo *n* , o vetor de peso atualizado $w_j(n+1)$ no tempo $n+1$ é definido por (KOHONEN, 1982; RITTER et al., 1992; KOHONEN, 1997a) :

$$w_j(n+1) = w_j(n) + \eta(n)h_{j,i(x)}(n)(x - w_j(n)) \qquad (6.8)$$

que é aplicado a todos os neurônios da grade que se encontram dentro da vizinhança topológica do neurônio vencedor *i* . A equação 6.8 tem o efeito de mover o peso sináptico w_i do neurônio vencedor *i* em direção ao vetor de entrada *x* . Através da apresentação repetida dos dados de treinamento, os vetores de peso sináptico tendem a seguir a distribuição dos vetores de entrada devido à atualização da vizinhança . O algoritmo, portanto, leva a uma ordenação topológica do mapa de características no espaço de entrada no sentido de que neurônios que são adjacentes na grade tenderão a ter vetores de peso sináptico similares.

A equação 6.8 é a formula desejada para calcular os pesos sinápticos do mapa de características. Além desta equação é preciso da heurística da equação 6.7 para selecionar a função de vizinhança $h_{j,i(x)}(n)$ e uma outra heurística para selecionar o parâmetro da taxa de aprendizagem $\eta(n)$.

O parâmetro da taxa de aprendizagem $\eta(n)$ deve ser variável no tempo como como indicado na equação 6.9, que corresponde ao caso da aproximação estocástica. Em particular, ele deve começar em um valor inicial η_0 e então decrescer gradualmente com o aumento do tempo n. Esta exigência pode ser satisfeita escolhendo-se um decaimento exponencial $\eta(n)$ como mostrado por

$$\eta(n) = \eta_0 \exp (- n / \tau_2), \text{ n} = 0,1,2,... \quad (6.9)$$

onde τ_2 é uma outra constante de tempo do algoritmo SOM. Apesar de as formulas de decaimento exponencial conforme equações 6.6 e 6.9 para a largura da função de vizinhança e o parâmetro da taxa de aprendizagem, respectivamente, poderem não ser ótimas, elas são normalmente adequadas para a formação do mapa de características de uma maneira auto-organizada.

6.1.5
As duas fases do processo adaptativo: ordenação e cooperação

Começando de um estado inicial de desordem completa, o algoritmo SOM gradualmente leva a uma representação organizada de padrões de ativação retirados do espaço de entrada, desde que os parâmetros do algoritmo sejam selecionados adequadamente. A adaptação de pesos sinápticos da grade pode ser decomposta em duas fases, uma fase de ordenação ou de auto-organização seguida por uma fase de convergência. Estas duas fases do processo adaptativo são descritas como segue (KOHONEN, 1982, 1997a) :

1. Fase de auto-organização ou de ordenação. É durante esta primeira fase do processo adaptativo que ocorre a ordenação topológica dos vetores de peso. A fase de ordenação pode exigir 1000 iterações do algoritmo SOM, e até mais. Deve-se levar em conta considerações sobre a escolha do parâmetro de aprendizagem e da função de vizinhança.

- O parâmetro da taxa de aprendizagem $\eta\ (n)$ deve iniciar com um valor próximo a 0,1, depois, ele deve decrescer gradualmente mas permanecer acima de 0,01. Estes valores desejáveis são satisfeitos pelas seguintes escolhas na formula da equação 6.9 :

$$\eta_0 = 0{,}1 \tag{6.10}$$

$$\tau_2 = 1000 \tag{6.11}$$

- A função de vizinhança $h_{j,i}\ (n)$ deve inicialmente incluir quase todos os neurônios da grade centrados no neurônio vencedor i , e então diminuir lentamente com o tempo. Especificamente, durante a fase de ordenação, que pode exigir 1000 iterações ou mais, permite-se que $h_{j,i}\ (n)$ se reduza a um valor pequeno de apenas um par de neurônios vizinhos em torno do neurônio vencedor ou ao próprio neurônio vencedor. Assumindo o uso de uma grade bidimensional de neurônios para o mapa discreto, pode-se igualar o tamanho inicial σ da função de vizinhança ao "raio" da grade. A constante de tempo τ_1 da formula 6.6 fica :

$$\tau_1 = 1000 / \log \sigma \tag{6.12}$$

2. Fase de convergência. Esta segunda fase do processo adaptativo é necessária para realizar uma sintonia fina do mapa de características e assim produzir uma quantização estatística precisa do espaço de entrada. Como regra geral, o número de interações que constituem a fase de convergência deve ser no mínimo 500 vezes o número de neurônios da grade. Assim, a fase de convergência pode durar milhares ou dezenas de milhares de iterações:

- Para uma boa precisão estatística, o parâmetro da taxa de aprendizagem $\eta\ (n)$ deve ser mantido durante a fase de convergência em um valor pequeno, da ordem de 0,01. Não deve permitir que ele diminua a zero

- A função de vizinhança $h_{j,i(x)}$ *(n)* deve conter apenas os vizinhos mais próximos de um neurônio vencedor, que pode eventualmente se reduzir a um ou a zero neurônios vizinhos.

6.2
Resumo do algoritmo SOM

A essência do algoritmo SOM de Kohonen é que ele substitui por uma computação geométrica simples as propriedades mais detalhadas da regra baseada em aprendizagem hebbiana e interações laterais. Os parâmetros essenciais do algoritmo são:

- Um espaço de entrada continuo de padrões de ativação que são gerados de acordo com uma certa distribuição de probabilidade
- Uma topologia da grade na forma de uma grade de neurônios, que define um espaço de saída discreto.
- Uma função de vizinhança variável no tempo $h_{j,i(x)}$ *(n)* que é definida em torno de um neurônio vencedor *i (x)*
- Um parâmetro da taxa de aprendizagem $\eta(n)$ que começa em um valor inicial η_0 e então diminui gradativamente com o tempo, *n*, mas nunca vai a zero.

Para a função de vizinhança e parâmetro da taxa de aprendizagem utiliza-se a equação 6.7 e 6.9, respectivamente, para a fase de ordenação (i.e., as primeiras mil iterações aproximadamente). Para uma boa precisão estatística, $\eta(n)$ deve ser mantido em um valor pequeno (0,01 ou menos) durante a convergência para um período de tempo razoavelmente longo, que é tipicamente de 1.000 iterações. Como no caso da função de vizinhança, ele deve conter apenas os vizinhos mais próximos do neurônio vencedor no início da fase de convergência e pode eventualmente diminuir a um ou a zero neurônios vizinhos.

Há três passos básicos envolvidos na aplicação do algoritmo após a inicialização: amostragem, casamento por similaridade e atualização. Estes três

passos são repetidos até a formação do mapa de características estar completa (HAYKIN, 2001) resumido como segue :

1. Inicialização . Escolher valores aleatórios para os vetores de peso iniciais $w_j(0)$. A única restrição aqui é que os $w_j(0)$ sejam diferentes para j = 1,2,...,l, onde l é o número de neurônios na grade. Um outro modo de inicializar o algoritmo é selecionar os vetores de peso $\{ w_j(0) \}^{l}_{j=1}$ a partir do conjunto disponível de vetores de entrada $\{ x_i \}^{N}_{i=1}$ de uma maneira aleatória.

2. Amostragem. Retirar uma amostra x do espaço de entrada com uma certa probabilidade. O vetor x representa o padrão de ativação que é aplicado à grade. A dimensão do vetor x é igual a m.

3. Casamento por similaridade. Encontrar o neurônio com o melhor casamento (neurônio vencedor) $i(x)$ no passo de tempo n usando o critério de mínima distancia euclidiana:

$$i(x) = \arg \min_j \| x(n) - w_j \|, \quad j = 1,2,...,l \tag{6.13}$$

4. Atualização. Ajuste os vetores de peso sináptico de todos os neurônios usando a formula de atualização

$$w_j (n + 1) = w_j (n) + \eta(n) h_{j,i(x)}(n)(x - w_j(n)) \tag{6.14}$$

onde $\eta(n)$ é parâmetro da taxa de aprendizagem de $h_{j,i(x)}(n)$ é a função de vizinhança centrada em torno do neurônio vencedor $i(x)$. Ambos $\eta(n)$ e $h_{j,i(x)}(n)$ são variados dinamicamente durante a aprendizagem para obter melhores resultados

5. Continuação. Continuar com o passo 2 até que não sejam observadas modificações significativas no mapa de características.