

FATORES DE HIPERTENSÃO USANDO REDES NEURAIS KOHONEN – SOM (SELF-ORGANIZING MAPS)

Autora: Nadia Maria Coelho Rodrigues

Aluna: Flaviane Peccin

Orientador: Prof. Ricardo Valgas

Introdução

- Rede Neural Artificial (RNA) é um modelo inspirado no funcionamento do cérebro com capacidade de adquirir e armazenar conhecimento.
- São utilizadas para : aproximação de funções, previsão de séries temporais, classificações, reconhecimento de padrões, extração de informações, suporte à decisão.

Introdução

- O Mapa Auto-organizável de Kohonen é um tipo especial de rede neural com aprendizado não-supervisionado.
- SOM – Kohonen fundamenta a formação de agrupamentos (ou *clusters*) de indivíduos de uma população, que apresentam comportamento semelhante em relação a determinadas características mensuradas.

Objetivo

- Extrair informações do banco de dados para reconhecimento de classes com características semelhantes.
- Encontrar associação entre as variáveis através da formação de agrupamentos (*clusters*).

Banco de dados

- Obtido através de inquérito domiciliar na Ilha do Governador – RJ com o objetivo de verificar a prevalência da hipertensão arterial na população adulta (maior de 20 anos).
- Amostragem por conglomerado em dois estágios estratificada pela classe socioeconômica.
- Entrevistados 1.270 indivíduos, moradores de 750 domicílios.
- Analisados no estudo 1.231 indivíduos (foram excluídos valores inválidos).

Variáveis

• Com o uso da regressão logística (ser ou não hipertenso) foram selecionadas para o estudo as seguintes variáveis:

- Idade
- Sexo
- Peso
- Altura
- Escolaridade
- Pressão arterial sistólica
- Pressão arterial diastólica
- Renda familiar
- Consumo de cigarros

Análise de Cluster

- Técnica exploratória de dados que tem por objetivo formar grupos de objetos semelhantes.
- Os clusters são obtidos por meio de aplicação dos conceitos de similaridade e distância.
- Utilizou-se a distância Euclidiana e a de Manhattan.

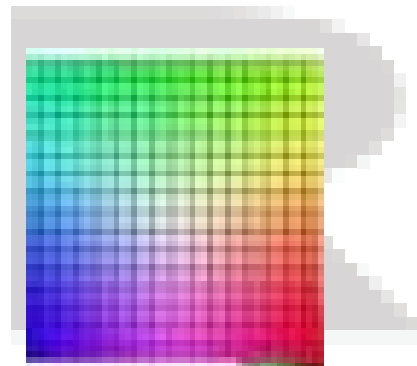
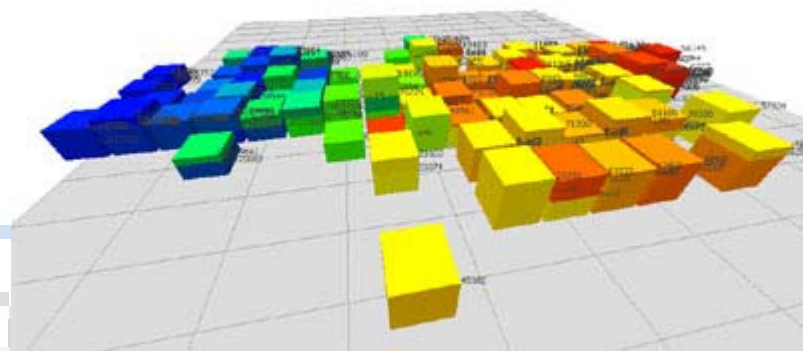
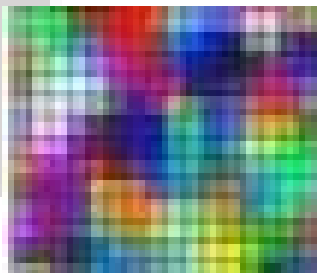
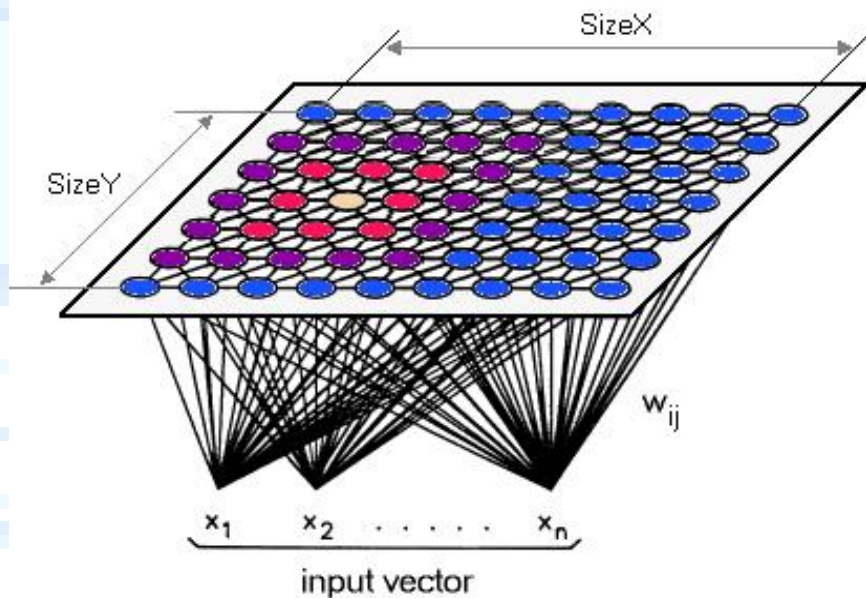
Euclidiana

$$d(\underline{x}, \underline{y}) = \sqrt{\sum_{i=1}^p (x_i - y_i)^2}$$

Manhattan

$$d(\underline{x}, \underline{y}) = \sum_{i=1}^p |x_i - y_i|$$

Algoritmo de Kohonen



Resultados

- Melhores resultados são encontrados suprimindo as variáveis sexo, escolaridade, renda e consumo de cigarros.
- As redes com menos de 500 vezes o número de processadores não apresentaram bons resultados.
- Através da análise das cinco características orgânicas dos indivíduos, obtém-se 3 grupos.

Resultados

	Idade	Cigarros	Peso	Altura	Sistólica	Diastólica
Cluster 1 (395)	40,57	16,28	77,56	170,81	135,76	84,17
Cluster 2 (475)	33,81	7,03	57,16	156,17	111,65	79,47
Cluster 3 (361)	58,21	7,11	62,01	151,92	151,52	88,24
Média Geral	43,54	9,51	66,41	162,59	131,12	81,12

* Os números entre parênteses indicam o número de indivíduos em cada cluster.

Conclusão

- SOM – *Kohonen* obteve uma excelente representação dos padrões usando uma rede 20x20, distância de Manhattan, 200 mil iterações para 5 variáveis, onde se visualizam 3 clusters.
- Podemos concluir que SOM – *Kohonen* pode ser considerada uma excelente técnica de *clusterização*, recomendando seu uso para mineração de dados.

Recomendações

- Estudar mais detalhadamente a eficiência do clusterizador.
- Fazer comparações com outros clusterizadores.

UFPR

UNIVERSIDADE FEDERAL DO PARANÁ