机器学习及其MATLAB实现—从基础到实践 第5课

法律声明

【声明】本视频和幻灯片为炼数成金网络课程的教学资料，所有资料只能在课程内使用，不得在课程以外范围散播，违者将可能被追究法律和经济责任。

课程详情访问炼数成金培训网站

http://edu.dataguru.cn

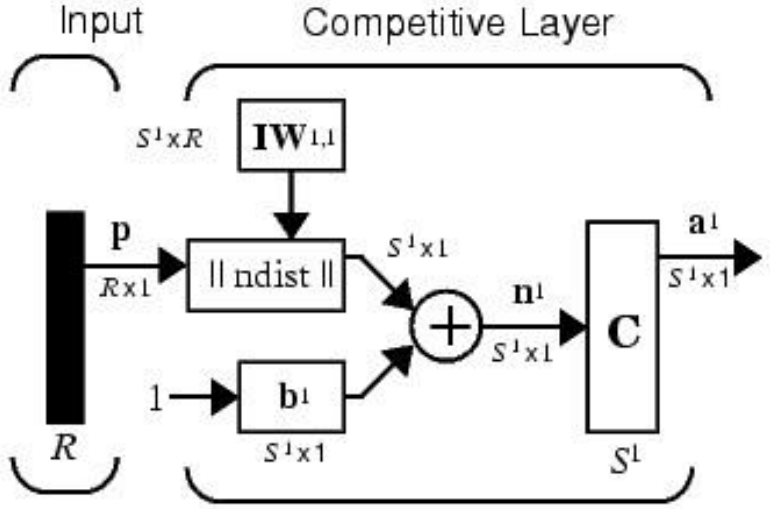
课程目录

Unsupervised Learning

- In machine learning, **unsupervised learning** is a class of problems in which one seeks to determine how the data are organized.
- It is distinguished from **supervised learning** (and **reinforcement learning**) in that the learner is given only **unlabeled** examples.
- Methods include:
  - ✓ clustering
  - ✓ blind source separation
  - ✓ self-organizing feature map
  - ✓ adaptive resonance theory
  - ✓ ……

竞争神经网络概述

- Competitive learning is useful for **classification** of input patterns into a discrete set of output classes.
- The neurons in a competitive layer distribute **themselves** to recognize frequently presented input vectors.
- Competitive learning is a rule based on the idea that **only one neuron** from a given iteration in a given layer will fire at a time.
- The "**winner**" of each iteration, element i^*, is the element whose total weighted input is the **largest**.
- Using this notation, one example of a competitive learning rule can be defined mathematically as:

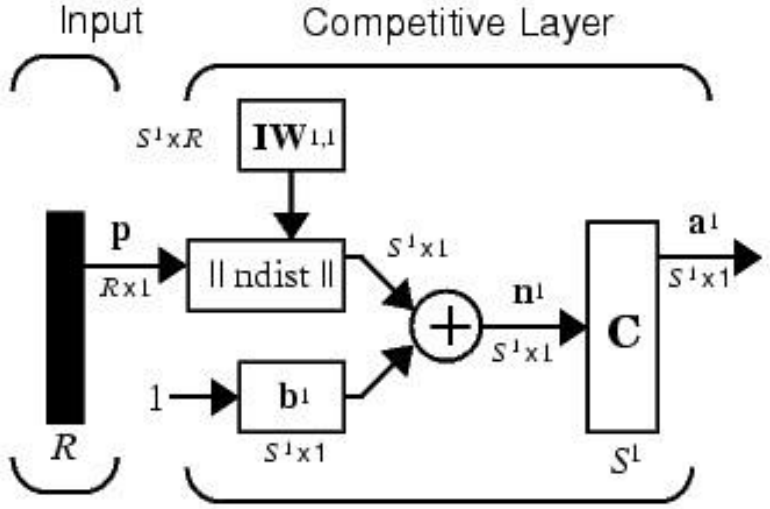


$$w_{ij}[n+1] = w_{ij}[n] + \Delta w_{ij}[n]$$

$$\Delta w_{ij}[n] = \begin{cases} \eta(x_i - w_{ij}) & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

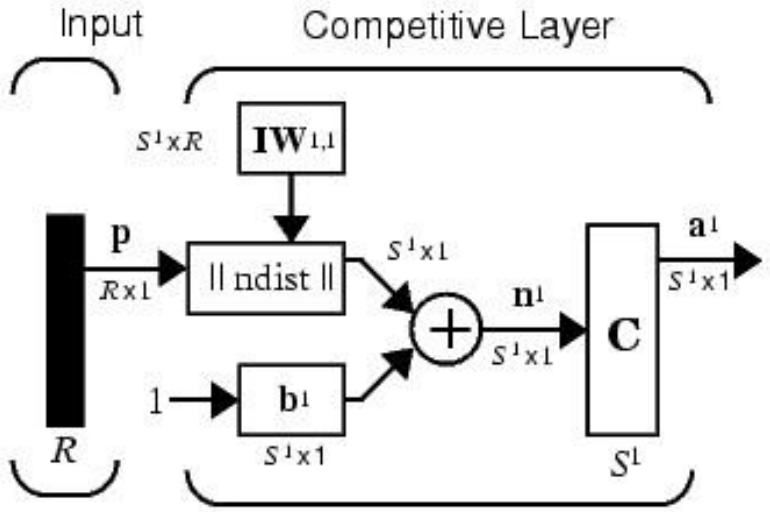
竞争神经网络概述

- The weights are initialized to the **centers** of the **input ranges** with the function **midpoint**.
- The biases are computed by **initcon**.
- Compute the net input $\mathbf{n}^1$ of a competitive layer by finding the **negative distance** between input vector **p** and the weight vectors and adding the biases **b**.
- If all biases are zero, the maximum net input a neuron can have is 0. This occurs when the input vector **p** equals that neuron's weight vector.
- The competitive transfer function accepts a net input vector for a layer and returns neuron outputs of 0 for all neurons **except for the *winner***, the neuron associated with the **most positive** element of net input n^1. The winner's output is 1.



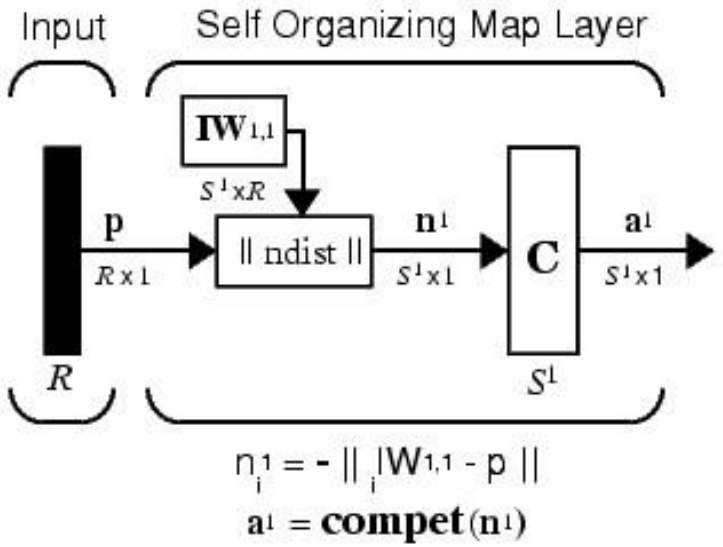
竞争神经网络概述

- The neuron whose weight vector was closest to the input vector **is updated to be even closer**. The result is that the winning neuron is **more likely to win** the competition **the next time** a similar vector is presented, and less likely to win when a very different input vector is presented.
- One of the **limitations** of competitive networks is that some neurons might **not always be *allocated***. These unfortunate neurons, referred to as ***dead neurons***, never perform a useful function.
- A **positive bias**, added to the negative distance, makes a distant neuron more likely to win.
- Update the biases with the learning function **learncon** so that the biases of **frequently** active neurons become **smaller**, and biases of **infrequently** active neurons become **larger**.



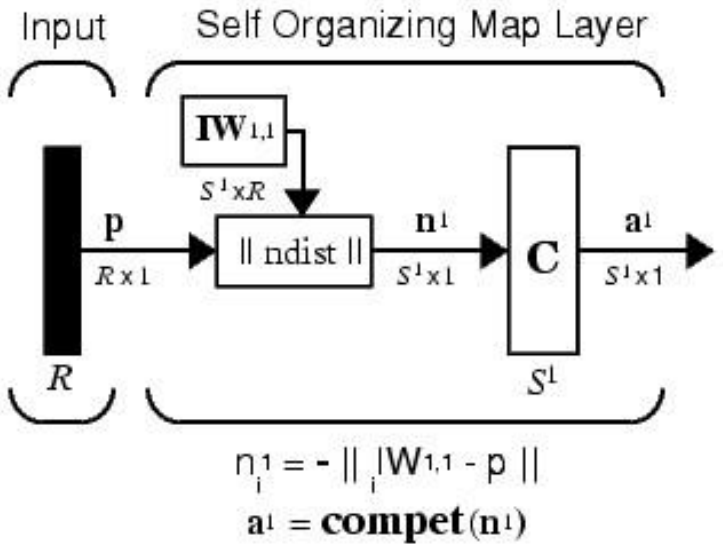
SOFM概述

- A **self-organizing map (SOM)** or **self-organizing feature map (SOFM)** is a type of artificial neural network that is trained using **unsupervised learning** to produce a low-dimensional (typically two-dimensional), discretized representation of the input space of the training samples, called a **map**.
- The self-organizing map describes **a mapping from a higher dimensional input space to a lower dimensional map space**.
- This architecture is like that of a competitive network, except **no bias** is used here.
- Instead of updating only the winning neuron, neurons **close to** the winning neuron are **updated along with** the winning neuron.



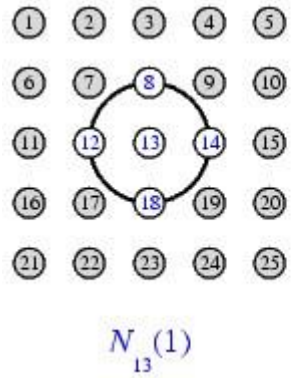
SOFM概述

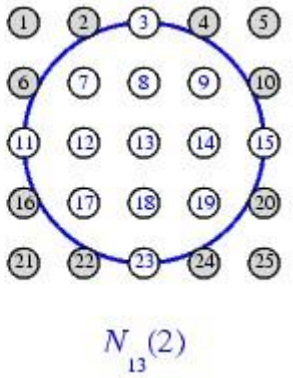
- The initialization for newsom is **midpoint**.
- The default learning in a self-organizing feature map occurs in the **batch mode** (**trainbu**). The weight learning function for the self-organizing map is **learnsomb**.
- The **distance** that defines the size of the neighborhood is **altered** during training through two phases.
  - ✓ **Ordering Phase**
  - ✓ **Tuning Phase**
- The neuron's weight vectors **initially take large steps** all together toward the area of input space , as the neighborhood size **decreases to 1**, the map tends to **order itself** topologically over the presented input vectors.

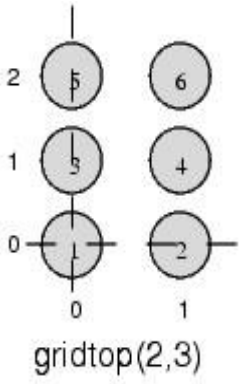


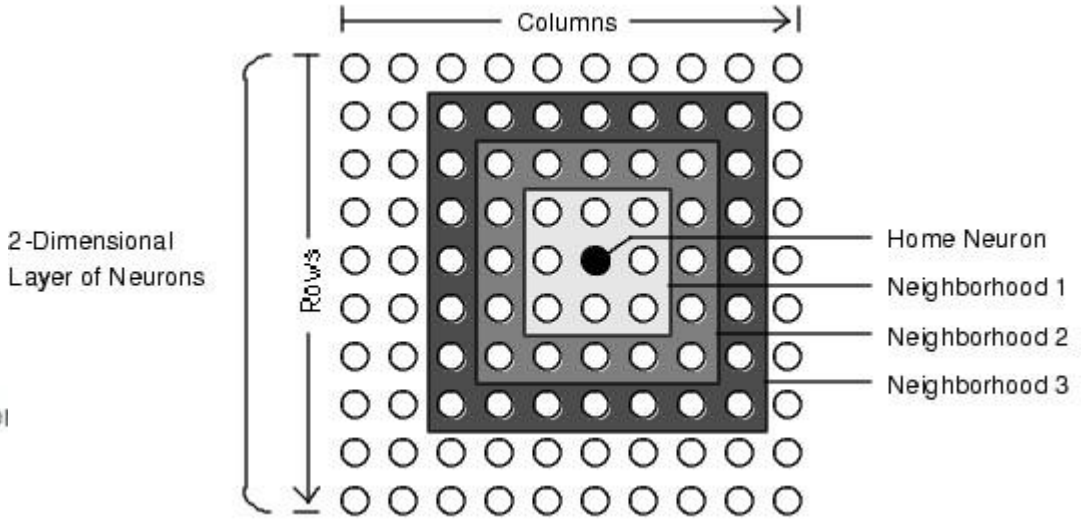
SOFM概述

- **Topologies**
  - ✓ gridtop
  - ✓ hextop
  - ✓ randtop
- **Distance Functions**
  - ✓ Dist
  - ✓ Linkdist
  - ✓ Mandist
  - ✓ boxdist
    - $_i\mathbf{w}(q) = {}_i\mathbf{w}(q-1) + \alpha(\mathbf{p}(q) - {}_i\mathbf{w}(q-1))$ 或
    - $_i\mathbf{w}(q) = (1-\alpha){}_i\mathbf{w}(q-1) + \alpha\mathbf{p}(q)$









重点函数解读

- **newc**
  - Create competitive layer
  - net = newc(PR,S,KLR,CLR)
- **newsom**
  - Create self-organizing map
  - net = newsom(P,[D1,D2,...],TFCN,DFCN,STEPS,IN)

案例分析

矿井突水水源判别

•相关研究表明，可以利用水化学法判别矿井的突水水源，其基本依据是：由于受到含水层的沉积期、地层岩性、建造和地化环境等诸多因素的影响，使储存在不同含水层中的地下水主要化学成分有所不同。

•为了准确地判别突水水源，需要综合多种因素，用的比较多的是“7大离子”溶解氧、硝酸根离子等。

$$Na^+、K^+、Ca^{2+}、Mg^{2+}、Cl^-、SO_4^{2-}和HCO_3^-$$

炼数成金逆向收费式网络课程

- Dataguru（炼数成金）是专业数据分析网站，提供教育，媒体，内容，社区，出版，数据分析业务等服务。我们的课程采用新兴的互联网教育形式，独创地发展了逆向收费式网络培训课程模式。既继承传统教育重学习氛围，重竞争压力的特点，同时又发挥互联网的威力打破时空限制，把天南地北志同道合的朋友组织在一起交流学习，使到原先孤立的学习个体组合成有组织的探索力量。并且把原先动辄成千上万的学习成本，直线下降至百元范围，造福大众。我们的目标是：低成本传播高价值知识，构架中国第一的网上知识流转阵地。
- 关于逆向收费式网络的详情，请看我们的培训网站 http://edu.dataguru.cn

Thanks

FAQ时间