机器学习及其MATLAB实现—从基础到实践 第7课

DATAGURU专业数据分析社区

法律声明

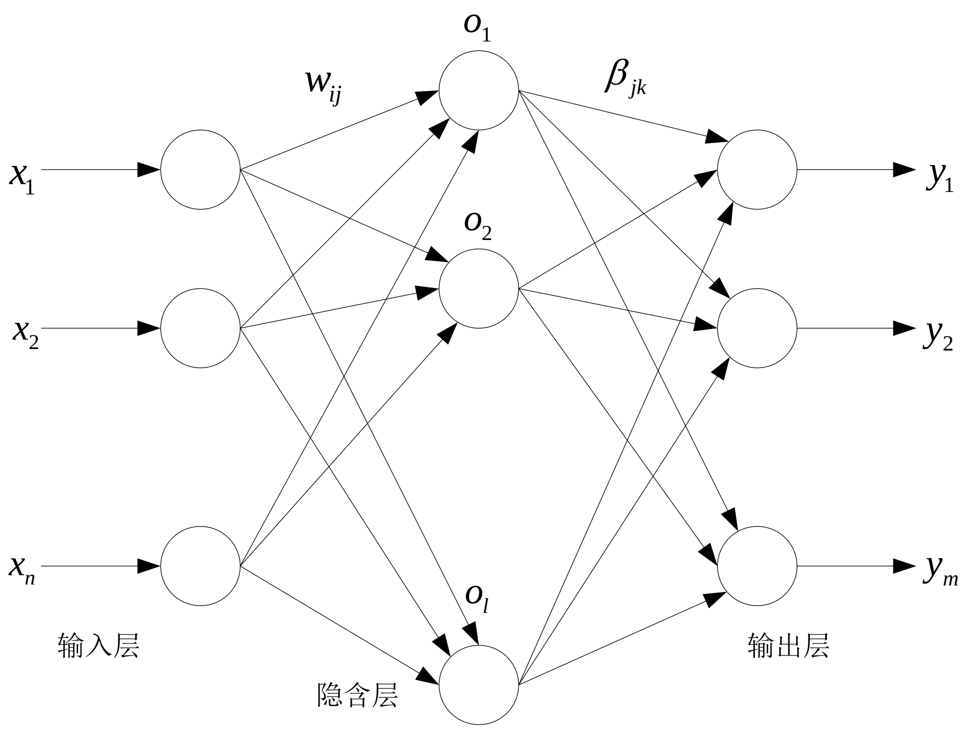
【声明】本视频和幻灯片为炼数成金网络课程的教学资料，所有资料只能在课程内使用，不得在课程以外范围散播，违者将可能被追究法律和经济责任。

课程目录

- 第一课　MATLAB入门基础
- 第二课　MATLAB进阶与提高
- 第三课　BP神经网络
- 第四课　RBF、GRNN和PNN神经网络
- 第五课　竞争神经网络与SOM神经网络
- 第六课　支持向量机（Support Vector Machine, SVM）
- **第七课　极限学习机（Extreme Learning Machine, ELM）**
- 第八课　决策树与随机森林
- 第九课　遗传算法（Genetic Algorithm, GA）
- 第十课　粒子群优化（Particle Swarm Optimization, PSO）算法
- 第十一课　蚁群算法（Ant Colony Algorithm, ACA）
- 第十二课　模拟退火算法（Simulated Annealing, SA）
- 第十三课　降维与特征选择

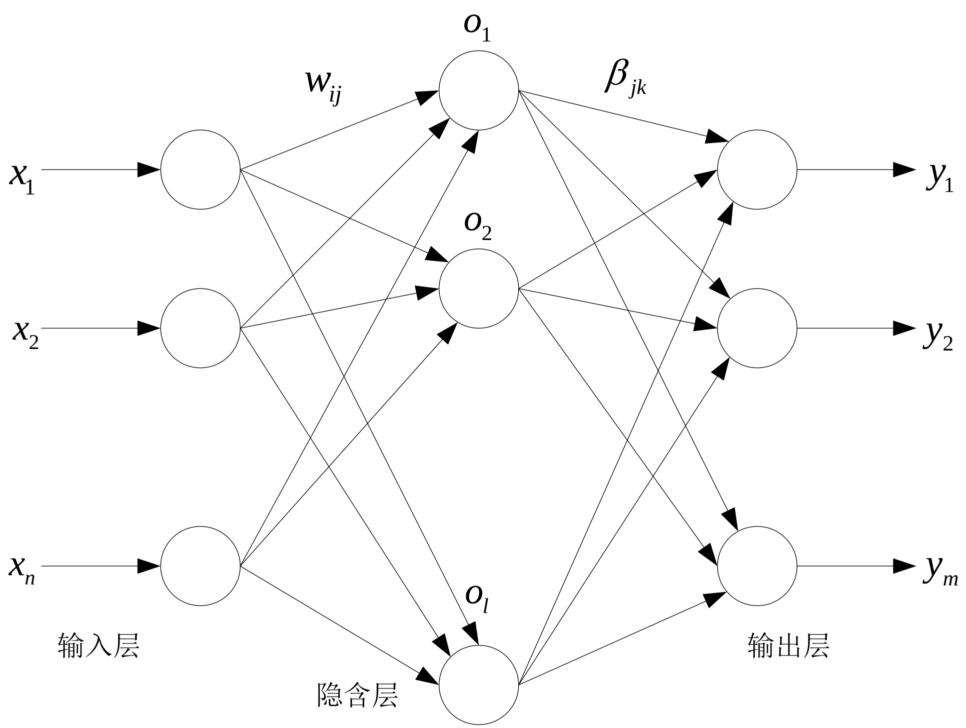
极限学习机原理概述

- The learning speed of feedforward neural networks is in general **far slower than required** and it has been a **major bottleneck** in their applications for past decades.
- Two key reasons behind may be:
  1) the slow **gradient-based** learning algorithms are extensively used to train neural networks.
  2) all the parameters of the networks are **tuned iteratively** by using such learning algorithms.



极限学习机原理概述

- Unlike these traditional implementations, **Extreme Learning Machine (ELM)** which **randomly** all the hidden nodes parameters of generalized **S**ingle-hidden **L**ayer **F**eedforward **N**etworks (**SLFN**s) and **analytically** determines the output weights of SLFNs.
- All the hidden node parameters are **independent** from the target functions or the training datasets. **All the parameters of ELMs can be analytically determined instead of being tuned.**
- In theory, this algorithm tends to provide the **good generalization performance** at extremely **fast learning speed**.

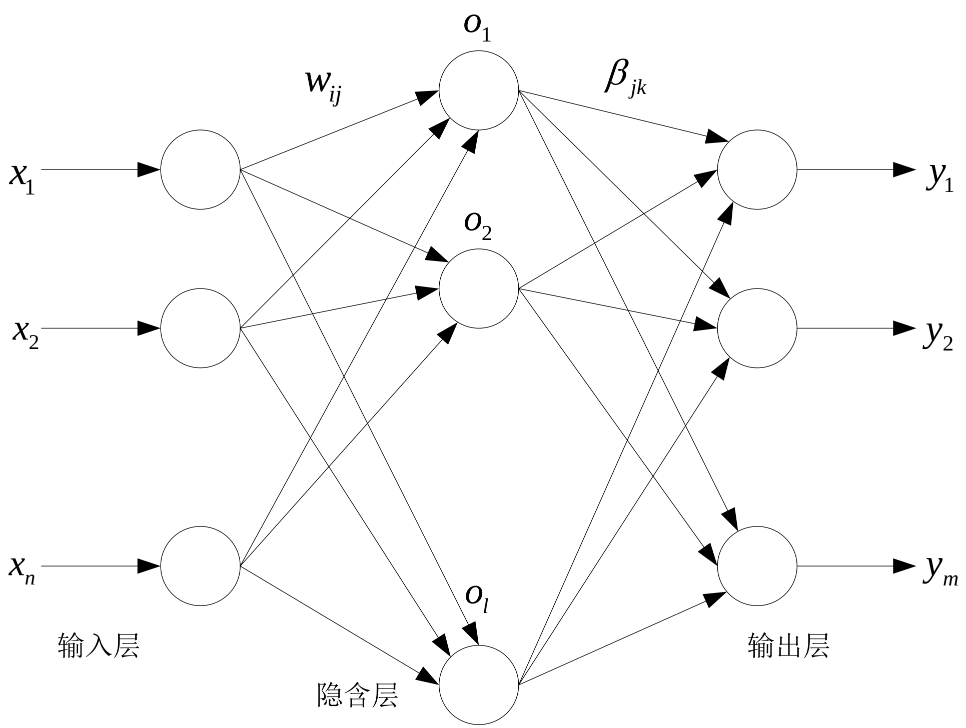


极限学习机原理概述

$$w=\begin{bmatrix} w_{11} & w_{12} & \dots & w_{1n} \\ w_{21} & w_{22} & \dots & w_{2n} \\ & \dots\dots & & \\ w_{l1} & w_{l2} & \dots & w_{ln} \end{bmatrix}_{l\times n} \qquad \beta=\begin{bmatrix} \beta_{11} & \beta_{12} & \dots & \beta_{1m} \\ \beta_{21} & \beta_{22} & \dots & \beta_{2m} \\ & \dots\dots & & \\ \beta_{l1} & \beta_{l2} & \dots & \beta_{lm} \end{bmatrix}_{l\times m} \qquad b=\begin{bmatrix} b_1 \\ b_2 \\ \dots \\ b_l \end{bmatrix}_{l\times 1}$$

$$X=\begin{bmatrix} x_{11} & x_{12} & \dots & x_{1Q} \\ x_{21} & x_{22} & \dots & x_{2Q} \\ & \dots\dots & & \\ x_{n1} & x_{n2} & \dots & x_{nQ} \end{bmatrix}_{n\times Q} \qquad Y=\begin{bmatrix} y_{11} & y_{12} & \dots & y_{1Q} \\ y_{21} & y_{22} & \dots & y_{2Q} \\ & \dots\dots & & \\ y_{m1} & y_{m2} & \dots & y_{mQ} \end{bmatrix}_{m\times Q}$$

$$T=\left[t_1,t_2,\dots,t_Q\right]_{m\times Q},\ t_j=\begin{bmatrix} t_{1j} \\ t_{2j} \\ \dots \\ t_{mj} \end{bmatrix}_{m\times 1}=\begin{bmatrix} \sum_{i=1}^{l}\beta_{i1}g\left(w_i x_j+b_i\right) \\ \sum_{i=1}^{l}\beta_{i2}g\left(w_i x_j+b_i\right) \\ \dots\dots \\ \sum_{i=1}^{l}\beta_{im}g\left(w_i x_j+b_i\right) \end{bmatrix}_{m\times 1} \quad \left(j=1,2,\dots,Q\right)$$

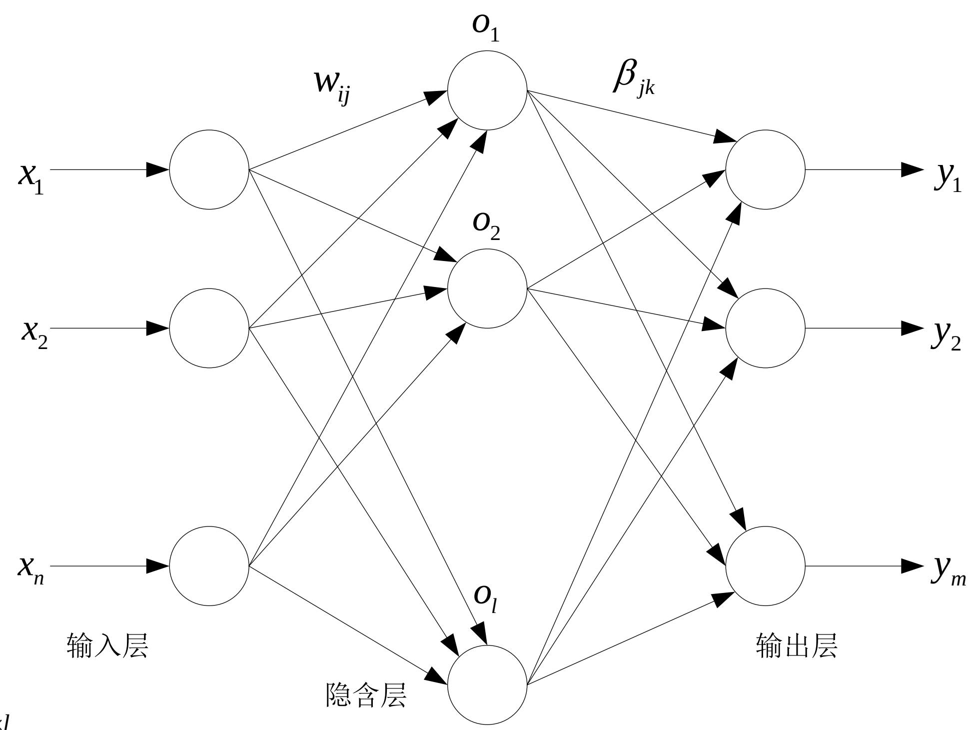


极限学习机原理概述

$$T=\left[t_1,t_2,\ldots,t_Q\right]_{m\times Q},\ t_j=\begin{bmatrix}t_{1j}\\t_{2j}\\\cdots\\t_{mj}\end{bmatrix}_{m\times 1}=\begin{bmatrix}\sum_{i=1}^{l}\beta_{i1}g\left(w_ix_j+b_i\right)\\\sum_{i=1}^{l}\beta_{i2}g\left(w_ix_j+b_i\right)\\\cdots\cdots\\\sum_{i=1}^{l}\beta_{im}g\left(w_ix_j+b_i\right)\end{bmatrix}_{m\times 1}\quad\left(j=1,2,\ldots,Q\right)$$

$$H\beta=T'$$

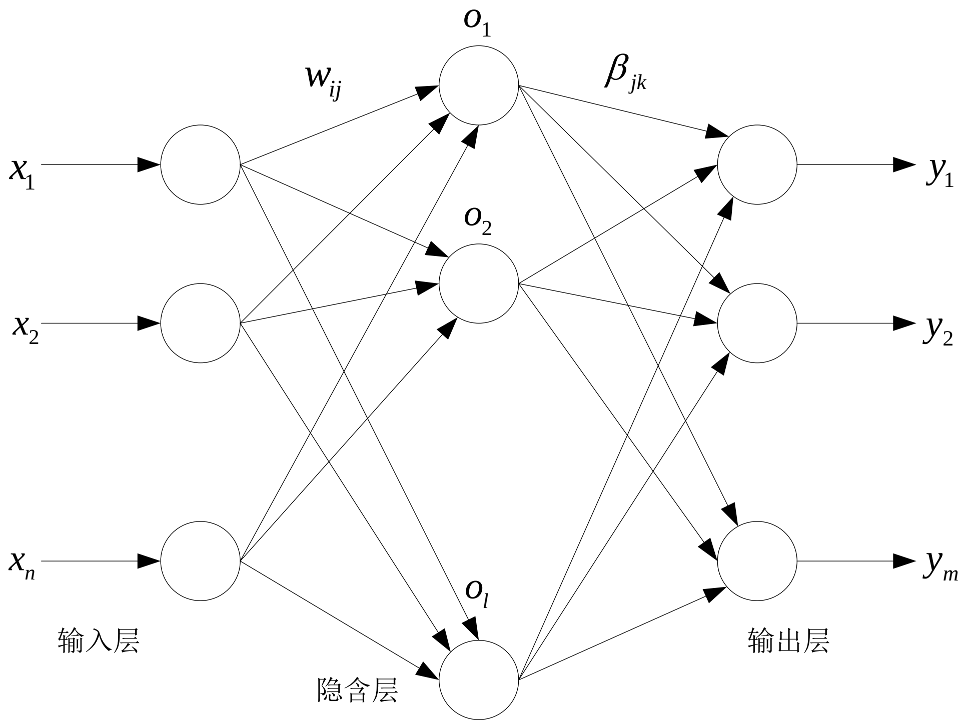
$$H\left(w_1,w_2,\ldots,w_l,b_1,b_2,\ldots,b_l,x_1,x_2,\ldots,x_Q\right)=\begin{bmatrix}g(w_1\cdot x_1+b_1) & g(w_2\cdot x_1+b_2) & g(w_l\cdot x_1+b_l)\\g(w_1\cdot x_2+b_1) & g(w_2\cdot x_2+b_2) & g(w_l\cdot x_2+b_l)\\ & \cdots\cdots & \\g(w_1\cdot x_Q+b_1) & g(w_2\cdot x_Q+b_2) & g(w_l\cdot x_Q+b_l)\end{bmatrix}_{Q\times l}$$



极限学习机原理概述

Theorem 2.1. *Given a standard SLFN with N hidden nodes and activation function $g: R \to R$ which is infinitely differentiable in any interval, for N arbitrary distinct samples $(\mathbf{x}_i, \mathbf{t}_i)$, where $\mathbf{x}_i \in \mathbf{R}^n$ and $\mathbf{t}_i \in \mathbf{R}^m$, for any $\mathbf{w}_i$ and b_i randomly chosen from any intervals of $\mathbf{R}^n$ and $\mathbf{R}$, respectively, according to any continuous probability distribution, then with probability one, the hidden layer output matrix $\mathbf{H}$ of the SLFN is invertible and $\|\mathbf{H}\beta - \mathbf{T}\| = 0$.*

Theorem 2.2. *Given any small positive value $\varepsilon > 0$ and activation function $g: R \to R$ which is infinitely differentiable in any interval, there exists $\tilde{N} \leqslant N$ such that for N arbitrary distinct samples $(\mathbf{x}_i, \mathbf{t}_i)$, where $\mathbf{x}_i \in \mathbf{R}^n$ and $\mathbf{t}_i \in \mathbf{R}^m$, for any $\mathbf{w}_i$ and b_i randomly chosen from any intervals of $\mathbf{R}^n$ and $\mathbf{R}$, respectively, according to any continuous probability distribution, then with probability one, $\|\mathbf{H}_{N\times\tilde{N}}\beta_{\tilde{N}\times m} - \mathbf{T}_{N\times m}\| < \varepsilon$.*



极限学习机原理概述

Algorithm ELM: Given a training set $\aleph = \{(\mathbf{x}_i, \mathbf{t}_i) | \mathbf{x}_i \in \mathbf{R}^n, \mathbf{t}_i \in \mathbf{R}^m, i = 1, \ldots, N\}$, activation function $g(x)$, and hidden node number $\tilde{N}$,

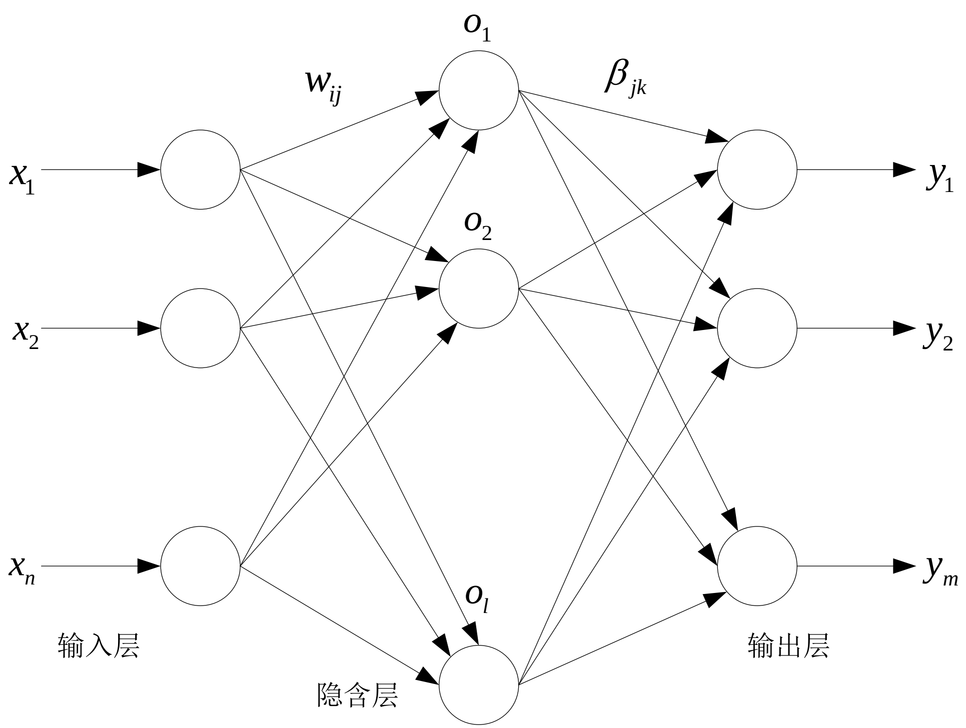
Step 1: Randomly assign input weight $\mathbf{w}_i$ and bias b_i, $i = 1, \ldots, \tilde{N}$.

Step 2: Calculate the hidden layer output matrix $\mathbf{H}$.

Step 3: Calculate the output weight β

$$\beta = \mathbf{H}^{\dagger}\mathbf{T}, \tag{16}$$

where $\mathbf{T} = [\mathbf{t}_1, \ldots, \mathbf{t}_N]^{\mathrm{T}}$.



极限学习机原理概述

- Compared BP Algorithm and SVM，ELM has several salient features:
  - **Ease of use.** No parameters need to be manually   tuned except predefined network architecture.
  - **Faster learning speed.** Most training can be completed in *milliseconds, seconds,* and *minutes*.
  - **Higher generalization performance.** It could obtain better generalization performance than BP in most cases, and reach generalization performance similar to or better than SVM.
  - **Suitable for almost all nonlinear activation functions.**. Almost all piecewise continuous (including discontinuous, differential, non-differential functions) can be used as activation functions.
  - **Suitable for fully complex activation functions**. Fully complex functions can also be used as activation functions in ELM.

重点函数解读

- nargin
- error
- pinv
- sin / hardlim
- elmtrain
- elmpredict

案例分析

汽油辛烷值预测

鸢尾花种类识别

案例分析

神经网络GUI实现

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