Homework 2

肖飞宇　2018210441

1 BOOSTING: FROM WEAK TO STRONG

1.1

Show that for each threshold s, there is some $m_0(s) \in \{0, 1, \cdots, m\}$ such that

$$\sum_{i=1}^{m} p_i 1\left\{\phi_{s,+}\left(x^{(i)}\right) \neq y^{(i)}\right\} = \frac{1}{2} - \frac{1}{2}\left(\sum_{i=1}^{m_0(s)} y^{(i)} p_i - \sum_{i=m_0(s)+1}^{m} y^{(i)} p_i\right) \tag{1.1}$$

and

$$\sum_{i=1}^{m} p_i \left\{\phi_{s,-}\left(x^{(i)}\right) \neq y^{(i)}\right\} = \frac{1}{2} - \frac{1}{2}\left(\sum_{i=m_0(s)+1}^{m} y^{(i)} p_i - \sum_{i=1}^{m_0(s)} y^{(i)} p_i\right) \tag{1.2}$$

Proof. 对于上述第一个等式，首先考虑到$x^{(i)}$为单调递减序列，给出

$$m_0(s) = \arg\min_i x^{(i)} \quad s.t. \quad x^{(i)} \geq s \tag{1.3}$$

即 $m_0(s)$ 为序列$\{x^{(i)}\}$的一个分界点，满足

$$\begin{cases} x^{(i)} \leq s & \text{if } i \leq m_0(s) \\ x^{(i)} \geq s & \text{if } i \geq m_0(s) \end{cases} \tag{1.4}$$

那么有

$$\sum_{i=1}^{m} p_i 1\left\{\phi_{s,+}\left(x^{(i)}\right) \neq y^{(i)}\right\} = \sum_{i=1}^{m_0(s)} p_i 1\left\{\phi_{s,+}\left(x^{(i)}\right) \neq y^{(i)}\right\} + \sum_{i=m_0(s)+1}^{m} p_i 1\left\{\phi_{s,+}\left(x^{(i)}\right) \neq y^{(i)}\right\} \tag{1.5}$$

$$= \sum_{i=1}^{m_0(s)} p_i 1\left\{y^{(i)} \neq -1\right\} + \sum_{i=m_0(s)+1}^{m} p_i 1\left\{y^{(i)} \neq 1\right\} \tag{1.6}$$

$$= \frac{1}{2}\sum_{i=1}^{m_0(s)} (1-y^{(i)})p^{(i)} + \frac{1}{2}\sum_{i=1}^{m_0(s)} (1+y^{(i)})p^{(i)} \tag{1.7}$$

$$= \frac{1}{2}\sum_{i=1}^{m} p^{(i)} - \frac{1}{2}\left(\sum_{i=1}^{m_0(s)} y^{(i)} p_i - \sum_{i=m_0(s)+1}^{m} y^{(i)} p_i\right) \tag{1.8}$$

$$= \frac{1}{2} - \frac{1}{2}\left(\sum_{i=1}^{m_0(s)} y^{(i)} p_i - \sum_{i=m_0(s)+1}^{m} y^{(i)} p_i\right) \tag{1.9}$$

对于上述的下式，类似可证 □

1.2

Define, for each $m_0 \in \{0, 1, \ldots, m\}$

$$f(m_0) = \sum_{i=1}^{m_0} y^{(i)} p_i - \sum_{i=m_0+1}^{m} y^{(i)} p_i \tag{1.10}$$

show that $\gamma = \frac{1}{2m}$ satisfies that

$$\max_{m_0} \left| f(m_0) \right| \geq 2\gamma \tag{1.11}$$

Proof.

$$\left| f(m_0) - f(m_0+1) \right| = \left| \sum_{i=1}^{m_0} y^{(i)} p_i - \sum_{m_0+1}^{m} y^{(i)} p_i - \sum_{i=1}^{m_0+1} y^{(i)} p_i + \sum_{m_0+2}^{m} y^{(i)} p_i \right| \tag{1.12}$$

$$= 2|y^{m_0+1} p_{m_0+1}| \tag{1.13}$$

$$=2|p_{m_0+1}| \tag{1.14}$$

所以我们可以得到一系列的等式

$$|f(m-1) - f(m)| = 2|p_m|$$

$$|f(m-2) - f(m-1)| = 2|p_{m-1}|$$

$$\cdots$$

$$|f(0) - f(1)| = 2|p_1|$$

上面各项相加可以得到

$$\sum_{i=1}^{m} |f(i-1) - f(i)| = 2\sum_{i=1}^{m} p_i = 2 \tag{1.15}$$

而

$$2 = \sum_{i=1}^{m} |f(i-1) - f(i)| \tag{1.16}$$

$$\le \sum_{i=1}^{m} (|f(i-1)| + |f(i)|) \tag{1.17}$$

$$\le \sum_{i=1}^{m} (\max_{m_0} \left| f(m_0) \right| + \max_{m_0} \left| f(m_0) \right|) \tag{1.18}$$

$$= 2m \max_{m_0} |f(m_0)| \tag{1.19}$$

即可以证明

$$\max_{m_0} |f(m_0)| \ge \frac{1}{m} = 2\gamma \tag{1.20}$$

□

1.3

Based on the above answer, how large margin γ can thresholded decision stumps guarantee on any training set $\left\{\left(x^{(i)}, y^{(i)}\right)\right\}_{i=1}^{m}$?
Give an upper bound on the number of thresholded decision stumps required to achieve zero error on a given training set.

Proof. 依据 Theorem 1 我们得到

$$J_t \le \sqrt{1-4\gamma^2} J_{t-1} \tag{1.21}$$

迭代到 t=1 有

$$J_t \le \left(\sqrt{1-4\gamma^2}\right)^t J_0 \tag{1.22}$$

注意到即使对于任意的数据集，我们总可以首先对于其 $\{x^{(i)}\}_{i=1}^{m}$进行一个排序（大致花费 $m \log m$的时间）使得其满足

$$x^{(k_1)} > x^{(k_2)} > \ldots > x^{(k_m)} \tag{1.23}$$

此时我们可以使用在 1.1 和 1.2中得到的结论，即

$$\sum_{i=1}^{m} p_i 1 \left\{ \phi_{s,+}\left(x^{(i)}\right) \neq y^{(i)} \right\} = \frac{1}{2} - \frac{1}{2} \left(\sum_{i=1}^{m_0(s)} y^{(i)} p_i - \sum_{i=m_0(s)+1}^{m} y^{(i)} p_i \right) \le \frac{1}{2} - \gamma \tag{1.24}$$

即有

$$J_0 \le \frac{1}{2} - \gamma \tag{1.25}$$

另外还可以至少确保弱分类器的 margin满足

$$\gamma = \frac{1}{2m} \tag{1.26}$$

代入有

$$J_t \le \left(\sqrt{1-4\gamma^2}\right)^t \left(\frac{1}{2} - \gamma\right) \tag{1.27}$$

如果需要实现 zero error，那么应有

$$J_t < \frac{1}{m} \tag{1.28}$$

求得

$$t = 2\frac{\log \frac{2}{m-1}}{\log 1 - \frac{1}{m^2}} \tag{1.29}$$

即为迭代次数的上界 □

2 Deep Neural Networks: Have a Try

2.1 1

最佳的设置见图2.1 采用最大规模的神经网络（$8 \times 8 \times 8 \times 8 \times 8$）具体的参数设置为：

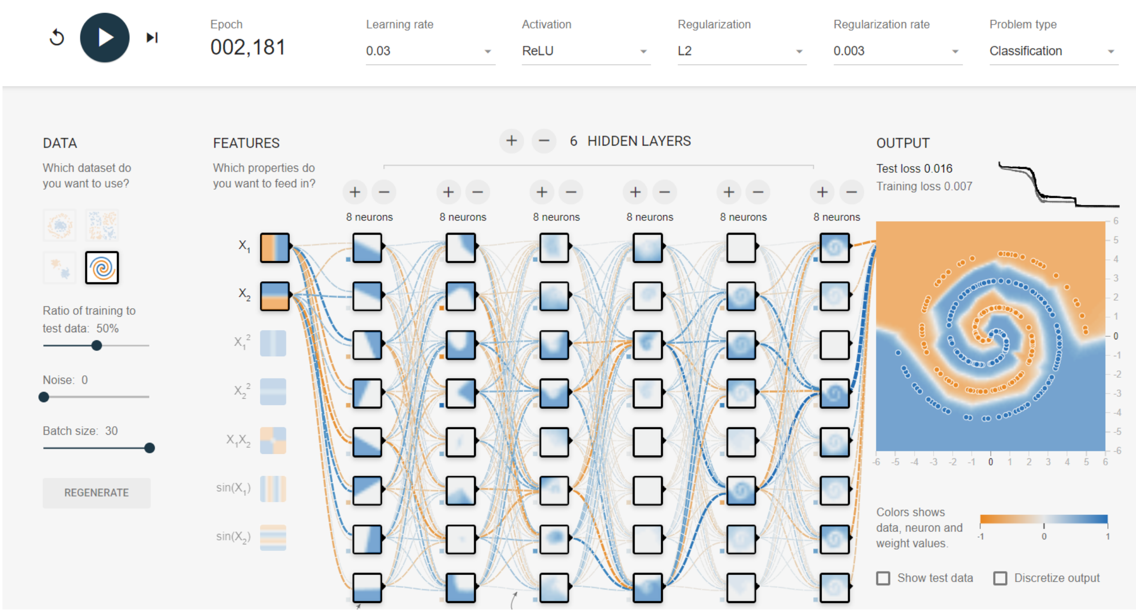

Figure 2.1: Configuration and results for the last dataset of Classification

- Learning rate 0.03
- Activation ReLu
- Regularization L_2
- Regularization rate 0.003

2.2 2

- Activation Function 选用ReLU激活函数效果最好
- Regularization L_2正则化效果比较好
- Hidden layers 此处采用层数较多的神经网络效果比较好（函数空间大）
- Learning rate, regularization rate 从图2.2中可以看出其的影响比较大

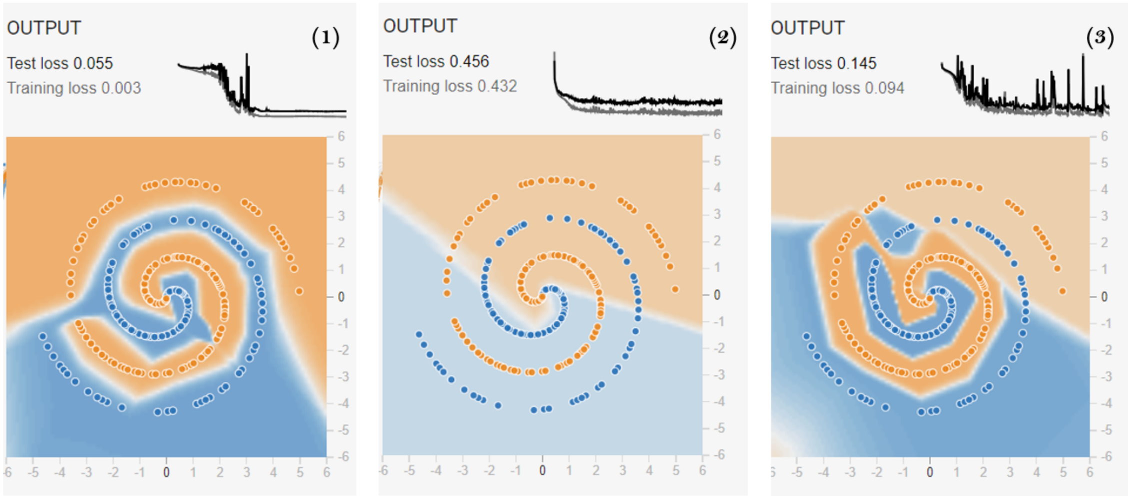

Figure 2.2: Choice of Learning rate and regularization rate:(1) learning rate = 0.1 (2)(3) regularization rate = 0.001 and 0.1

总而言之，ReLU 激活函数在这一问题中的收敛比较快，采用更多的隐层可以使得其收敛性变好（在恰当的正则化条件下），但是会使得收敛变慢。特别地，相对较小的学习率虽然使得收敛变慢，但是却表现更好（学习率过大引起剧烈震荡），正则化参数的选择也对于模型的收敛有关键的作用。

2.3 3

- 首先需要依据问题本身的复杂度估计一个恰当的网络特性，如其层数，还有其激活函数
- 注意对于学习率和正则系数的恰当选取，控制好误差
- 需要多次尝试，不要希望很快寻找到最优的参数

3 Clustering: Mixture of Multinomials

3.1 MLE for multinomial

Derive the maximum-likelihood estimator for the parameter $\boldsymbol{\mu} = (\mu_i)_{i=1}^{d}$ of a multinomial distribution:

$$P(\boldsymbol{x}|\boldsymbol{\mu}) = \frac{n!}{\prod_i x_i!}\prod_i \mu_i^{x_i}, \quad i = 1, \cdots, d \tag{3.1}$$

where $x_i \in \mathbb{N}, \sum_i x_i = n$ and $0 < \mu_i < 1, \sum_i \mu_i = 1$.

Proof. 直接取似然函数为 $P(\boldsymbol{x}|\boldsymbol{\mu})$，即

$$L(\boldsymbol{\mu}) == \frac{n!}{\prod_i x_i!}\prod_i \mu_i^{x_i} = n!\prod_i \frac{\mu_i^{x_i}}{x_i!} \tag{3.2}$$

约束为

$$\sum_i \mu_i = 1 \tag{3.3}$$

拉格朗日函数为

$$LR(\boldsymbol{\mu}, \lambda) = n! \prod_i \frac{\mu_i^{x_i}}{x_i!} - \lambda \sum_i \mu_i \tag{3.4}$$

对 μ_i 的导数为零，有

$$\frac{\partial}{\partial \mu_i} L(\boldsymbol{\mu}) - \lambda \frac{\partial}{\partial \mu_i} \sum_i \mu_i = 0 \tag{3.5}$$

即

$$\frac{x_i}{\mu_i} L(\boldsymbol{\mu}) = \lambda \tag{3.6}$$

即 x_i和μ_i成比例，于是设

$$\mu_i = k x_i \tag{3.7}$$

累加并代入 $\sum_i \mu_i = 1$和$\sum_i x_i = n$求得

$$\mu_i = \frac{x_i}{n} \tag{3.8}$$

即求得对参数 μ_i 的最大似然估计 □

3.2 EM for mixture of multinomials

使用对数似然函数进行分析，首先注意到

$$P(d) = \sum_{k=1}^{K} P(d|c_d = k) P(c_d = k) = \frac{n_d!}{\prod_w T_{dw}!} \sum_{k=1}^{K} \pi_k \prod_w \mu_{wk}^{T_{dw}} \tag{3.9}$$

其中 n_d和T_{dw}均和参数无关所以在最大似然估计中可以简化为可以简化为

$$P(d) = \Delta \sum_{k=1}^{K} \pi_k \prod_w \mu_{wk}^{T_{dw}} \tag{3.10}$$

其中 Δ 为一个常数，那么对数似然函数可以简化为

$$MLP(\pi, \mu) = \sum_{d=1}^{D} \log \left(\sum_{k=1}^{K} \pi_k \prod_w \mu_{wk}^{T_{dw}} \right) \tag{3.11}$$

对于参数的约束为

$$\sum_{k=1}^{K} \pi_k = 1 \tag{3.12}$$

$$\sum_{w=1}^{W} \mu_{wk} = 1 \quad k \in [1, K] \tag{3.13}$$

则拉格朗日函数为

$$L(\pi, \mu, \alpha, \beta) = \sum_{d=1}^{D} \log \left(\sum_{k=1}^{K} \pi_k \prod_w \mu_{wk}^{T_{dw}} \right) + \alpha \left(1 - \sum_{k=1}^{K} \pi_k \right) + \sum_{k=1}^{K} \beta_k \left(1 - \sum_{w=1}^{W} \mu_{wk} \right) \tag{3.14}$$

分别对 π_k 和 μ_{wk}求导得到

$$\frac{\partial L}{\partial \pi_k} = \sum_{d=1}^{D} \frac{\prod_w \mu_{wk}^{T_{dw}}}{\sum_j \pi_j \prod_i \mu_{ij}^{T_{di}}} - \alpha \tag{3.15}$$

$$\frac{\partial L}{\partial \mu_{wk}} = \sum_{d=1}^{D} \frac{\pi_k \left(\prod_{i\neq w} \mu_{ik}^{T_{di}}\right) T_{dw} \mu_{wk}^{T_{dw}-1}}{\sum_j \pi_j \prod_i \mu_{ij}^{T_{di}}} - \beta_k \tag{3.16}$$

上述导数为0得到

$$\alpha = \sum_w \sum_d \frac{\prod_w \mu_{wk}^{T_{dw}}}{\sum_j \pi_j \prod_i \mu_{ij}^{T_{di}}} \tag{3.17}$$

$$\beta_k = \sum_w \sum_d \frac{\prod_w \mu_{wk}^{T_{dw}}}{\sum_j \pi_j \prod_i \mu_{ij}^{T_{di}}} T_{dw} \tag{3.18}$$

显然有耦合系数为

$$\gamma(c_{dk}) = \frac{\pi_k \prod_w \mu_{wk}^{T_{dw}}}{\sum_j \pi_j \prod_w \mu_{wj}^{T_{dw}}} \tag{3.19}$$

从而方程3.17和3.18可以改写为

$$\alpha = \sum_w \sum_d \gamma(c_{dk}) \tag{3.20}$$

$$\beta_k = \sum_w \sum_d \gamma(c_{dk}) T_{dw} \tag{3.21}$$

同时有

$$\pi_k = \frac{\sum_d \gamma(c_{dk})}{\sum_k \sum_d \gamma_{c_{dk}}} \tag{3.22}$$

$$\mu_{uk} = \frac{\sum_d \gamma(c_{dk}) T_{dw}}{\sum_w \sum_d \gamma(c_{dk}) T_{dw}} \tag{3.23}$$

那么EM算法中E步骤就是利用方程3.19估计γ的值，而M步就是利用方程3.22和3.23估计参数π和μ的值

3.3 算法

具体的 EM 算法的实现见 https://github.com/feiyuxiaoThu/SML/blob/master/Assigenments/Assignment2-EM-Algorithm/Multinomial_EM.py.

3.4 结果

随着选择的话题数目的增多，训练时间变长

不同话题数目下的高频词见下述的列表，总结而言，K=5和K=20是比较好的选择，其原因如下:

1. 对于K=5的情况，可以从深色的单词看出基本可以完成一个粗略的文本划分，其各自的主题对应也比较明显
2. 对于K=20的情况，可以看出是一个比较细致的划分

K=5

- Topic 0
  model network function algorithm learning data set neural result parameter system number error distribution input problem vector point weight neuron
- Topic 1
  network learning input training function neural set unit algorithm error data weight model output system problem number result hidden vector
- Topic 2
  network model input learning neuron system neural function output set cell result data pattern visual signal algorithm unit weight information
- Topic 3
  network learning algorithm unit model neural function set input problem result number training method output data weight system point action
- Topic 4
  model data learning network function input set system neural parameter training point problem result number algorithm neuron method object error

K=10

- Topic 0
  model network learning data input set function system neural training algorithm neuron result number cell output problem method error layer
- Topic 1
  network input model neuron cell neural unit pattern system learning function firing output activation set dynamic noise synaptic activity layer
- Topic 2
  model learning function network algorithm neural system action set method parameter problem distribution result data point number task input field
- Topic 3
  model algorithm function learning data set vector input problem method feature cell number result space features network training policy kernel
- Topic 4
  network algorithm problem unit vector function system set representation training neural output weight input hand data result parameter layer object
- Topic 5
  network model neural neuron input learning function set weight unit system training problem output algorithm error result word recognition point
- Topic 6
  network learning function model input neural algorithm unit set training data weight

error output system problem result number parameter hidden

- Topic 7

  model image data algorithm system cell signal set result images output function motion input neural network point recognition field circuit
- Topic 8

  network input model neuron system neural cell signal output information circuit result noise data function visual set unit analog learning
- Topic 9

  network object learning input set model system image features point memory task training neural unit result pattern view algorithm representation

K=20

- Topic 0

  network input neural system function algorithm unit model output layer training data learning weight set recognition result error number problem
- Topic 1

  network model neural neuron function set unit input cell object system output data signal point view learning problem result training
- Topic 2

  network system unit learning set model data training algorithm output neuron word neural target control hidden method result speech motion
- Topic 3

  neuron model cell network input function neural learning weight system set data visual error result training output direction number signal
- Topic 4

  learning weight set error network function algorithm training model data classifier neural problem system test result pattern input generalization cell
- Topic 5

  network function neural model neuron input set result data cell weight circuit pattern unit synaptic point system number training learning
- Topic 6

  network model system neural map unit cortex visual cell neuron position direction data object pattern field brain space place head
- Topic 7

  learning model network data set student input unit training error weight output number hidden function vector point teacher order parameter
- Topic 8

  network model data training set learning neural system input unit function error algorithm output weight number vector result net parameter

- Topic 9
  data label feature unit recognition model learning algorithm vector pattern training image performance set facial images class action classifier hidden
- Topic 10
  network learning model algorithm function input set data neuron parameter neural training system error problem result linear method vector number
- Topic 11
  input cell neuron model information unit motion function activity point direction result output visual rate light field orientation parameter equation
- Topic 12
  network model learning neural unit training set weight function input algorithm problem error data number output method vector result system
- Topic 13
  model learning weight network function algorithm data result input system neural problem set parameter method number space training probability noise
- Topic 14
  model system network input set vector auditory output unit point data neural cell training sound signal function learning problem component
- Topic 15
  network model learning input function neural algorithm system unit output neuron problem result data weight set pattern parameter signal number
- Topic 16
  model input learning function network cell system visual error output direction set neural data position phase weight eye vector circuit
- Topic 17
  network neuron object layer input vector circuit model output neural system cell visual function algorithm problem tree hand shape information
- Topic 18
  model system network input neural cell field control function noise dynamic information result distribution visual motion parameter word output set
- Topic 19
  network function learning set algorithm training data input error model problem system neural method number result output weight point space

K=30

- Topic 0
  network function model input neuron neural cell system point orientation unit visual connection result field data output dynamic set layer
- Topic 1

model network data learning function set neural input system algorithm neuron problem training cell result method information number unit point

- Topic 2

  network model training neural set input learning output unit error system data function weight algorithm parameter problem information layer object

- Topic 3

  network learning function algorithm model neural input weight set system unit problem result error training number output action method pattern

- Topic 4

  model input network cell output neural neuron weight data set system parameter function number distribution frequency vector current unit field

- Topic 5

  learning network model function input system training set error algorithm unit weight data neural output problem result parameter pattern hidden

- Topic 6

  network set neural model training algorithm neuron function result method problem learning input system weight information number performance data unit

- Topic 7

  network system cell neuron neural model correlation distribution mean frequency firing result response temporal burst eeg activity phase rate unit

- Topic 8

  input network learning pattern data unit set component system output neuron layer neural signal training linear information function point model

- Topic 9

  network system data input model neural set output training signal learning image speech result problem function analysis images method unit

- Topic 10

  cell input function model network neuron noise synaptic system output direction result neural rate layer learning distribution response pattern head

- Topic 11

  network data neural set training input error model vector performance cost representation features number current subscriber problem map word neuron

- Topic 12

  network learning function algorithm model unit input problem data weight output vector result number layer error set local training system

- Topic 13

  network neural model system set data result point input function method neuron analog theorem noise net error probability parameter nonlinear

- Topic 14

network neural algorithm training set problem learning result point classifier input data system output number unit function method vector error

- Topic 15

  network learning input algorithm data function training neural set vector weight number unit problem error result output space system matrix

- Topic 16

  model set data input function point unit image motion analog number system direction object algorithm probability noise result distribution error

- Topic 17

  network function input neural learning model weight output system neuron recognition layer set net number training task result unit pattern

- Topic 18

  learning classifier system algorithm model network neural set control function speaker training error data result vector recognition speech parameter hmm

- Topic 19

  algorithm function model network problem vector learning set neural error method weight point result solution number system data parameter input

- Topic 20

  network learning unit hidden weight error training generalization function algorithm model parameter data term set student noise vector number large

- Topic 21

  model data function learning algorithm network parameter error set training result problem neural number component distribution linear likelihood system input

- Topic 22

  network unit input neural system training output learning data model function error set weight algorithm hidden problem information vector number

- Topic 23

  layer network function cell input learning erp set response component eeg result parameter single data disparity problem weight point light

- Topic 24

  learning network algorithm model function unit system weight point vector input number parameter part result neural feature set data information

- Topic 25

  network model function neuron input unit system output neural training set number learning word hidden result context pattern cell recognition

- Topic 26

  set network model algorithm training learning unit examples neural object cell distribution pattern result frequency neuron visual system number position

- Topic 27

learning distribution function algorithm data model input error point optimal loss examples order case linear perceptron spike result output motion

- Topic 28

  input neuron model algorithm learning function set basis data problem network contrast output vector firing orientation partition visual smo training

- Topic 29

  model network neural control input cell forward system learning function output controller result feedback data inverse dynamic parameter training information