

データマイニング

第6回 自己組織化マップ

2023年春学期

宮津和弘

本日の講義・演習

日付	講義・演習内容
04/14/23	(1) イントロダクション
04/21/23	(2) ビジネスシミュレーション
04/28/23	(3) ID-POSデータ分析
05/12/23	(4) 対応分析
05/19/23	(5) クラスター分析
05/26/23	(6) 自己組織化マップ
06/02/23	(7) 線形判別分析
06/09/23	(8) 非線形判別分析
06/16/23	(9) ツリーモデル
06/23/23	(10) 集団学習
06/30/23	(11) サポートベクターマシン
07/04/23	(12) ネットワーク分析
07/14/23	(13) 共分散構造分析
07/21/23	(14) テキスト分析
07/28/23	(15) まとめ



本日の演習概要とポイント

- **ニューラルネットワークの基礎**
 - アルゴリズムと迷惑メールへの応用例
- **自己組織化マップ^o**
 - ニューラルネットワークを活用した手法





ニューラルネットワークの基礎



ニューラルネットワークとは

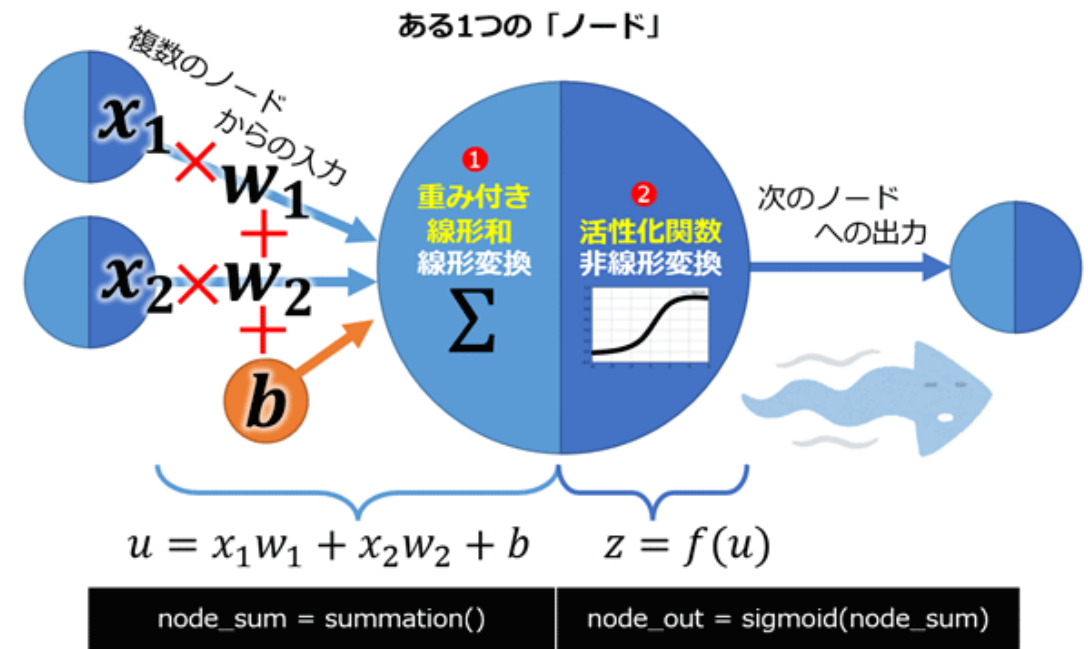
生物の神経回路を構成する最小単位である神経細胞（ニューロン）を数理的にモデル化したものが、**ニューラルネットワーク**である。

各樹形突起（ノード）からの入力は、シナプスの重みで乗されて加算される。

$$u = x_1w_1 + x_2w_2 + b$$

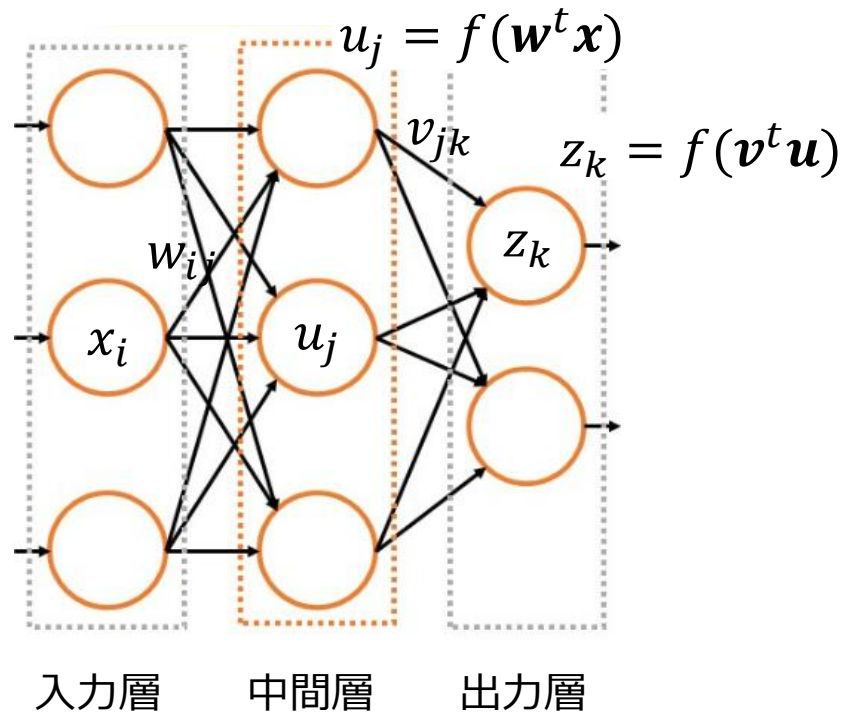
活性化関数に従って、次ノードの入力として出力される。

$$z = f(u)$$

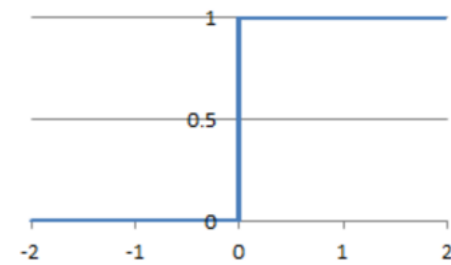


ニューラルネットワークの原理

出力信号と現実信号の差が最小となるように各層の重みを決定する

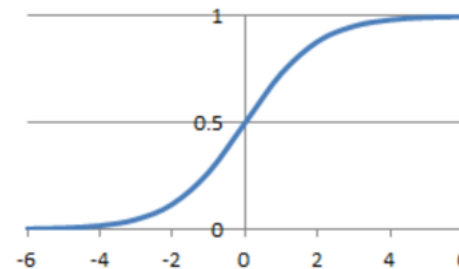


ステップ関数



$$f(t) = \begin{cases} 0 & (t < 0) \\ 1 & (t \geq 0) \end{cases}$$

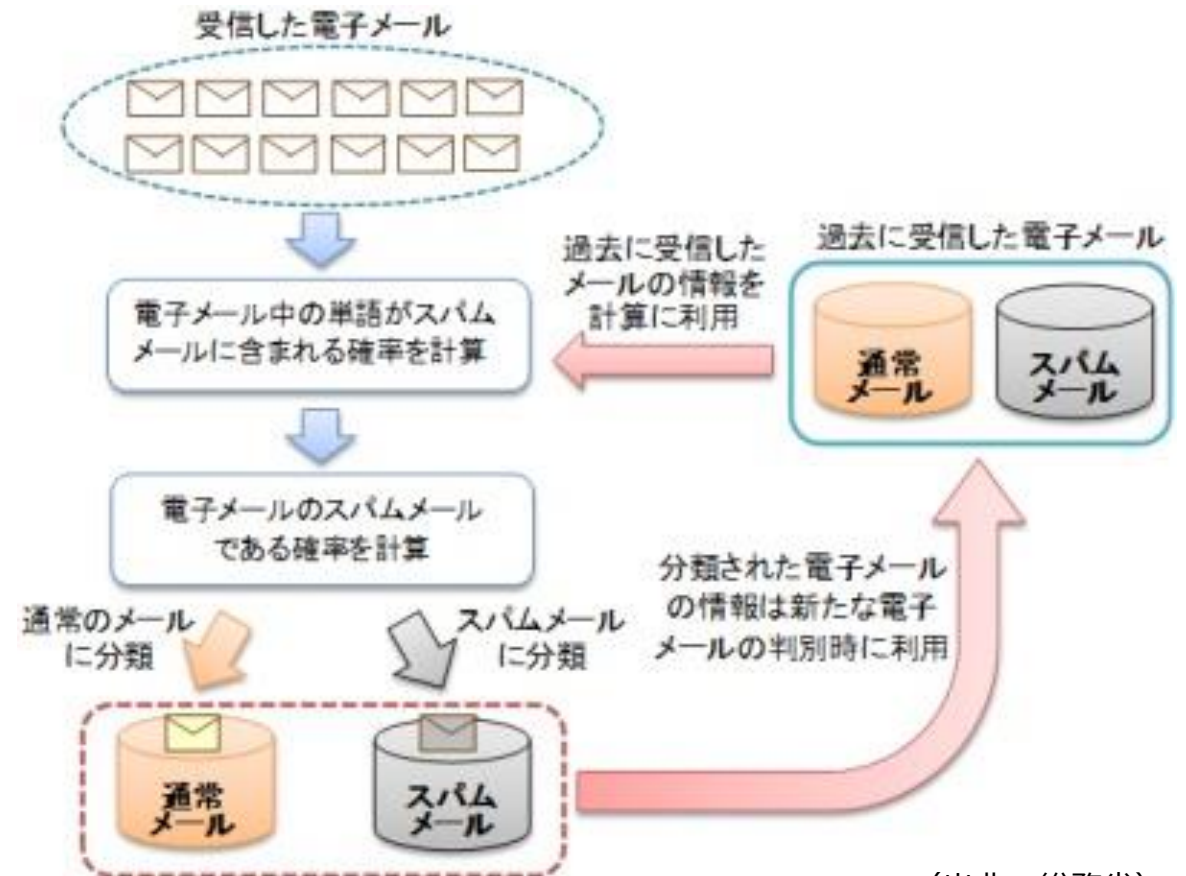
シグモイド関数



$$f(t) = \frac{1}{1 + e^{-t}}$$

ニューラルネットワークによるスパムメールへの応用

最近ではメールシステムにおいて、**スパムメール**（“迷惑メール”）をフィルタリング機能により、効率的かつ自動的に排除している。ある受信メールが、スパムまたは通常のメールなのかを判別するために、メールのタイトルや本文中の単語の出現する確率によって両者を判別するデータサイエンス手法が活用されている。



（出典：総務省）

スパムメールの語源

イギリスのコメディ番組の中で、レストランに入った夫婦が注文しようとする、店員に「スパム、スパム、スパム……」と連呼されてしまい、仕方なくスパムを注文することになってしまったという内容のコントがあった。その結果、迷惑な行為を「スパム」と呼ぶようになってしまったらしい…。



迷惑メール

VS.



通常メール

判別用のSPAMデータセット

SMSメッセージが**SPAM**か**HAM**かを内容とラベルで提供
このデータを用いて、ニューラルネットワークで学習して判別に用いる！

v1	v2
ham	Go until jurong point, crazy.. Available only in bugis n great world la e buffet... Cine there got amore wat...
ham	Ok lar... Joking wif u oni...
spam	Free entry in 2 a wkly comp to win FA Cup final tkts 21st May 2005. Text FA to 87121 to receive entry question(std txt rate)T&C's apply 08452810075over18's
ham	U dun say so early hor... U c already then say...
ham	Nah I don't think he goes to usf, he lives around here though
spam	FreeMsg Hey there darling it's been 3 week's now and no word back! I'd like some fun you up for it still? Tb ok! XxX std chgs to send, 螢1.50 to rcv
ham	Even my brother is not like to speak with me. They treat me like aids patent.
ham	As per your request 'Melle Melle (Oru Minnaminunginte Nurungu Vettam)' has been set as your callertune for all Callers. Press *9 to copy your friends Callertune
spam	WINNER!! As a valued network customer you have been selected to receivea 螢900 prize reward! To claim call 09061701461. Claim code KL341. Valid 12 hours only.
spam	Had your mobile 11 months or more? U R entitled to Update to the latest colour mobiles with camera for Free! Call The Mobile Update Co FREE on 08002986030
ham	I'm gonna be home soon and i don't want to talk about this stuff anymore tonight, k? I've cried enough today.
spam	SIX chances to win CASH! From 100 to 20,000 pounds txt> CSH11 and send to 87575. Cost 150p/day, 6days, 16+ TsandCs apply Reply HL 4 info

↑
“スパムメール”のフラグ

5,574サンプル

因子分析モデルのパス図

実際のSMSメッセージを57項目の特徴量で定量化したデータを入力とする

```
> head(spam)
  make address  all num3d  our over remove internet order mail receive will people report addresses free business email you credit your font num000
1  0.00    0.64  0.64    0  0.32  0.00    0.00    0.00  0.00  0.00    0.00  0.64  0.00  0.00    0.00  0.32    0.00  1.29  1.93    0.00  0.96    0    0.00
2  0.21    0.28  0.50    0  0.14  0.28    0.21    0.07  0.00  0.94    0.21  0.79  0.65  0.21    0.14  0.14    0.07  0.28  3.47    0.00  1.59    0    0.43
3  0.06    0.00  0.71    0  1.23  0.19    0.19    0.12  0.64  0.25    0.38  0.45  0.12  0.00    1.75  0.06    0.06  1.03  1.36    0.32  0.51    0    1.16
4  0.00    0.00  0.00    0  0.63  0.00    0.31    0.63  0.31  0.63    0.31  0.31  0.31  0.00    0.00  0.31    0.00  0.00  3.18    0.00  0.31    0    0.00
5  0.00    0.00  0.00    0  0.63  0.00    0.31    0.63  0.31  0.63    0.31  0.31  0.31  0.00    0.00  0.31    0.00  0.00  3.18    0.00  0.31    0    0.00
6  0.00    0.00  0.00    0  1.85  0.00    0.00    1.85  0.00  0.00    0.00  0.00  0.00  0.00    0.00  0.00    0.00  0.00  0.00    0.00  0.00    0    0.00

  money hp hp1 george num650 lab labs telnet num857 data num415 num85 technology num1999 parts pm direct cs meeting original project re edu table
1  0.00  0  0    0    0  0  0  0    0  0  0  0  0  0.00  0  0  0.00  0  0  0.00  0  0.00  0  0.00  0.00  0
2  0.43  0  0    0    0  0  0  0    0  0  0  0  0  0.07  0  0  0.00  0  0  0.00  0  0.00  0  0.00  0.00  0
3  0.06  0  0    0    0  0  0  0    0  0  0  0  0  0.00  0  0  0.06  0  0  0.12  0  0.06  0.06  0.06  0
4  0.00  0  0    0    0  0  0  0    0  0  0  0  0  0.00  0  0  0.00  0  0  0.00  0  0.00  0.00  0.00  0
5  0.00  0  0    0    0  0  0  0    0  0  0  0  0  0.00  0  0  0.00  0  0  0.00  0  0.00  0.00  0.00  0
6  0.00  0  0    0    0  0  0  0    0  0  0  0  0  0.00  0  0  0.00  0  0  0.00  0  0.00  0.00  0.00  0

 conference charsemicolon charRoundbracket charsquarebracket charExclamation charDollar charHash capitalAve capitalLong capitalTotal type
1          0          0.00          0.000          0          0.778          0.000          0.000          3.756          61          278 spam
2          0          0.00          0.132          0          0.372          0.180          0.048          5.114          101         1028 spam
3          0          0.01          0.143          0          0.276          0.184          0.010          9.821          485         2259 spam
4          0          0.00          0.137          0          0.137          0.000          0.000          3.537          40          191 spam
5          0          0.00          0.135          0          0.135          0.000          0.000          3.537          40          191 spam
6          0          0.00          0.223          0          0.000          0.000          0.000          3.000          15           54 spam
```

“スパムメール”のフラグ



特徴量について

Feature	Feature description	Feature criteria
#1	Message Length (ML)	The number of characters in the message
#2	Number of Words (NW)	The number of words in message, usually spam messages contains large number of words
#3	Ratio of Number of Words with length less than three (RNW3)	Number of words less than 3 (NW3) over to the total number of words (NW)
#4	Ratio of Number of Capital (RCW)	Number of Capital Words (CW) over total number of words (NW)
#5	Ratio of Alphanumeric Characters (RAC)	Number of Alphanumeric Characters over Message Length (ML)
#6	Ratio of Special Characters (RSC) such as “*, _ ,+,%,\$,@, , \/, ”	Number of Special Characters over Message Length (ML)
#7	Ratio of Punctuation Characters (RPC) such as “; ? ! : () – “ « » < > [] { } ”	Number of Punctuation Characters over Message Length (ML)
#8	Total number of Digit Characters (DC)	Normalized by maximum number of digit characters.
#9	The existence of word “Call” and digits together	The value of this field is either true or false
#10	The existence of URL in the message	The value of this field is either true or false

ニューラルネットワークの“h2o”パッケージを実行する

```
> library(kernlab)
> data(spam)
> install.packages("h2o")
> h2o.init()
> spam.hex <- as.h2o(spam)
> tr <- h2o.deeplearning(x=1:57,y=58,training_frame=spam.hex,hidden=c(30,20,10),epochs=100,nfolds=3)
```

※ 途中でアクセス権限でログイン等を求められるが無視して進める！

The logo for H2O.ai, featuring the text "H2O.ai" in a bold, black, sans-serif font on a solid yellow square background.

H2O's Deep Learning is based on a multi-layer [feedforward artificial neural network](#) that is trained with stochastic gradient descent using back-propagation. The network can contain a large number of hidden layers consisting of neurons with tanh, rectifier, and maxout activation functions. Advanced features such as adaptive learning rate, rate annealing, momentum training, dropout, L1 or L2 regularization, checkpointing, and grid search enable high predictive accuracy. Each compute node trains a copy of the global model parameters on its local data with multi-threading (asynchronously) and contributes periodically to the global model via model averaging across the network.

H2Oニューラルネットワーク分析の結果

```
> tr
Model Details:
=====
```

```
H2OBinomialModel: deeplearning
Model ID: DeepLearning_model_R_1684737511175_20
Status of Neuron Layers: predicting type, 2-class classification, bernoulli distribution, CrossEntropy loss, 2,592 weights/biases,
layer units      type dropout      l1      l2 mean_rate rate_rms momentum mean_weight weight_rms mean_bias bias_rms
1      1      57      Input  0.00 %      NA      NA      NA      NA      NA      NA      NA      NA      NA
2      2      30 Rectifier 0.00 % 0.000000 0.000000 0.007093 0.006915 0.000000 0.016980 0.246196 0.489926 0.136121
3      3      20 Rectifier 0.00 % 0.000000 0.000000 0.002000 0.002087 0.000000 -0.016872 0.241578 0.996793 0.108739
4      4      10 Rectifier 0.00 % 0.000000 0.000000 0.052070 0.201006 0.000000 0.001897 0.337078 1.005759 0.077393
5      5       2   Softmax      NA 0.000000 0.000000 0.082615 0.237244 0.000000 0.048575 1.589730 -0.002022 0.136087
```

```
MSE: 0.01868405
RMSE: 0.1366896
LogLoss: 0.074299
Mean Per-Class Error: 0.02418944
AUC: 0.9969718
AUCPR: 0.9957151
Gini: 0.9939436
```

混同行列

Confusion Matrix (vertical: actual; across: predicted) for F1-optimal threshold:

	nonspam	spam	Error	Rate
nonspam	2750	38	0.013630	=38/2788
spam	63	1750	0.034749	=63/1813
Totals	2813	1788	0.021952	=101/4601

全体の誤り率は2%程度

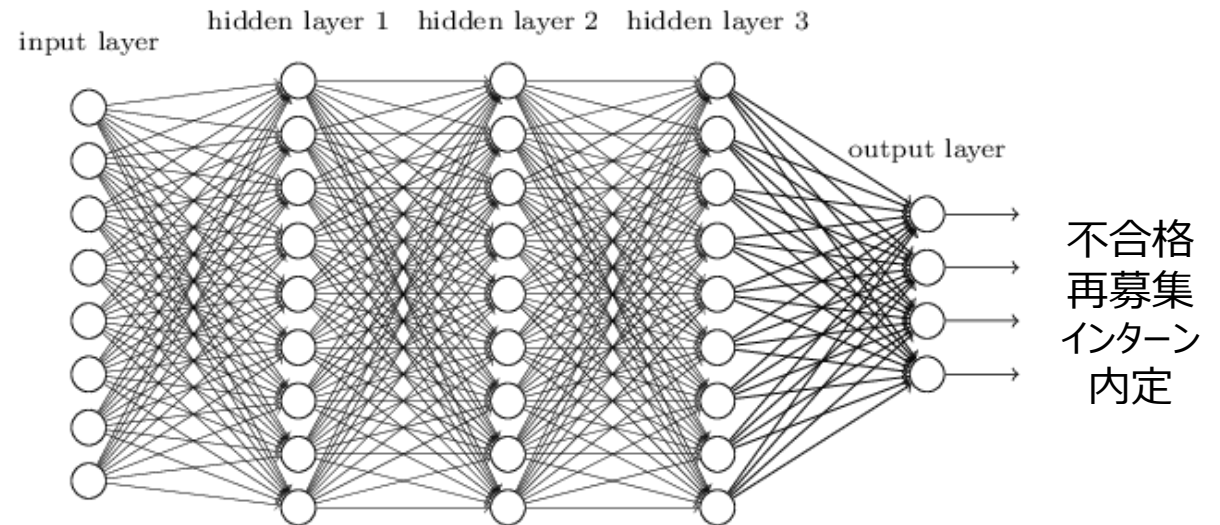
エントリーシートをAIが評価する（再掲）

エントリーシート

フリガナ	
氏名	
住所	
連絡先	

学生時代に力を入れたこと

私は学生時代、経営学の研究とその発表に注力しました。将来経営に深く携わりたいと考えた私は、経営学ゼミでリーダーシップについて学び、報告会参加を目指しました。最も困難だった点は経営を自分の言葉で説明することです。実際に経営の経験をしていないということに原因があると考え、以下の2つの事を行いました。一つ目はインプットの質向上です。概説書だけでなく、実際に経営者が受ける動画講習を見ることで、具体性を上げました。二つ目は経営者へのヒアリングです。自分がリサーチを通じて生まれた意見や仮説を、実際の経営者にぶつけることで検証しました。結果、学部生初の研究発表を行い、現場を見に行く行動力の大切さを学びました。

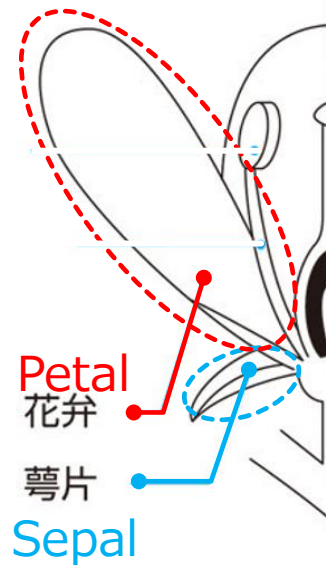


“ソフトバンクは、2017年からIBM Watsonを活用して
AIエントリーシート選考により75%時間削減した”

自己組織化マップ

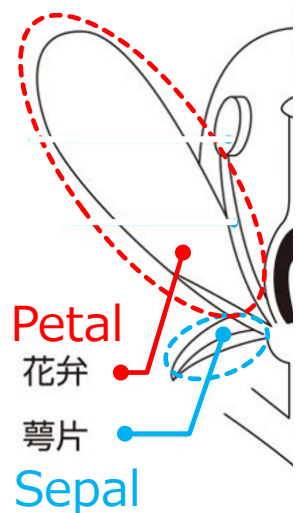
Rでよく使われるIRISデータについて

Rでよく使われるIRISデータとは、3種類のアヤメ（Versicolor、Virginica、Setosa）について花弁（Petal）とがく片（Sepal）の幅と長さについてサンプルデータ



IRISデータについて

アヤメの種類によって、
花弁と**がく片**の幅と長さ
で分類するためのデータ
として用いられる



Sepal.Length	Sepal.width	Petal.Length	Petal.width	Species
5.1	3.5	1.4	0.2	setosa
4.9	3.0	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa
4.6	3.1	1.5	0.2	setosa
5.0	3.6	1.4	0.2	setosa
5.4	3.9	1.7	0.4	setosa
4.6	3.4	1.4	0.3	setosa
5.0	3.4	1.5	0.2	setosa
4.4	2.9	1.4	0.2	setosa
4.9	3.1	1.5	0.1	setosa
7.0	3.2	4.7	1.4	versicolor
6.4	3.2	4.5	1.5	versicolor
6.9	3.1	4.9	1.5	versicolor
5.5	2.3	4.0	1.3	versicolor
6.5	2.8	4.6	1.5	versicolor
5.7	2.8	4.5	1.3	versicolor
6.3	3.3	4.7	1.6	versicolor
4.9	2.4	3.3	1.0	versicolor
6.6	2.9	4.6	1.3	versicolor
5.2	2.7	3.9	1.4	versicolor
6.3	3.3	6.0	2.5	virginica
5.8	2.7	5.1	1.9	virginica
7.1	3.0	5.9	2.1	virginica
6.3	2.9	5.6	1.8	virginica
6.5	3.0	5.8	2.2	virginica
7.6	3.0	6.6	2.1	virginica
4.9	2.5	4.5	1.7	virginica
7.3	2.9	6.3	1.8	virginica
6.7	2.5	5.8	1.8	virginica
7.2	3.6	6.1	2.5	virginica



試しに、クラスター分析してみよう

Rコマンダーから 【データ】 → 【パッケージ内のデータ】 → 【アタッチされたパッケージからデータセットを…】

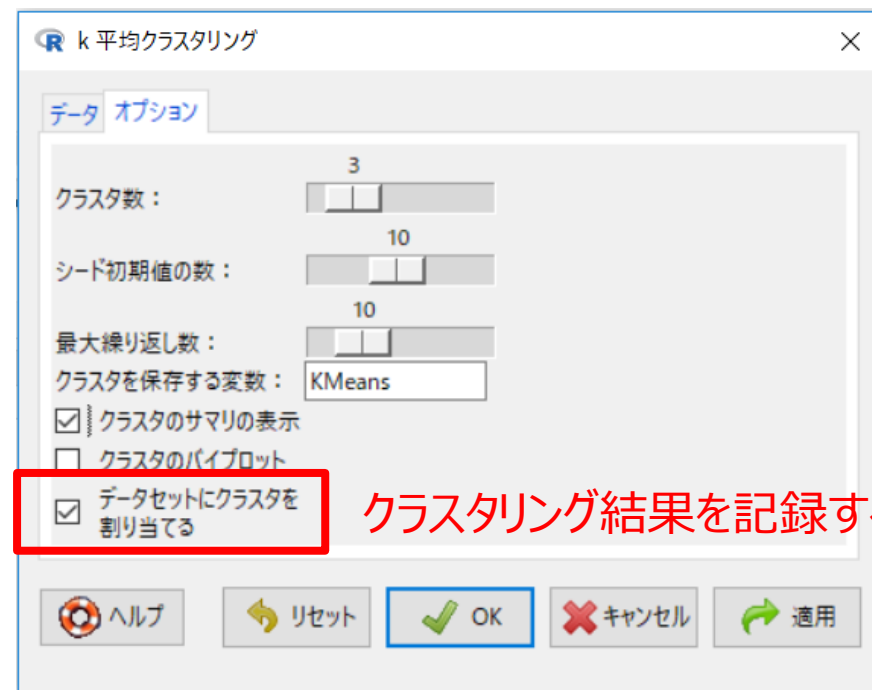
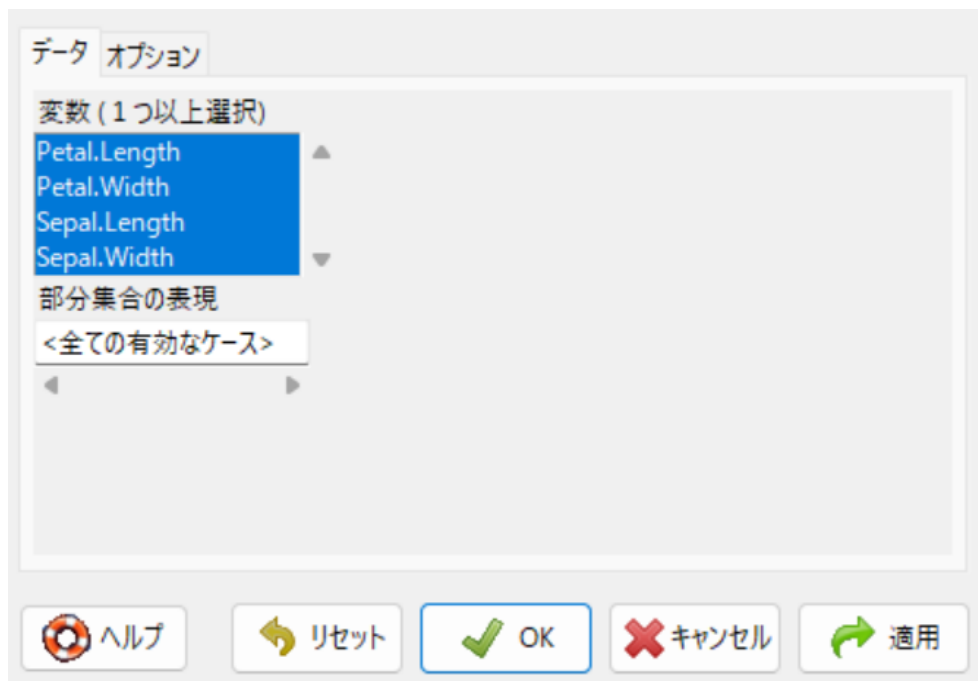


データセット名に“**iris**”入力してOKすると、
irisデータが読み込まれるのを確認（↓）

		1 2 3 4 5				
	rowname	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
1	1	5.1	3.5	1.4	0.2	setosa
2	2	4.9	3.0	1.4	0.2	setosa
3	3	4.7	3.2	1.3	0.2	setosa
4	4	4.6	3.1	1.5	0.2	setosa
5	5	5.0	3.6	1.4	0.2	setosa
6	6	5.4	3.9	1.7	0.4	setosa
7	7	4.6	3.4	1.4	0.3	setosa
8	8	5.0	3.4	1.5	0.2	setosa
9	9	4.4	2.9	1.4	0.2	setosa
10	10	4.9	3.1	1.5	0.1	setosa
11	11	5.4	3.7	1.5	0.2	setosa
12	12	4.8	3.4	1.6	0.2	setosa

非階層的クラスタリングの推定：K平均法

Rコマンダーから【統計量】→【次元解析】→【クラスター分析】→【k-平均クラスター分析】



クラスタリング結果を記録する！

アヤメに関する4つの特徴量を選択

K平均法のクラスタリング結果

rowname	1	2	3	4	5	6
	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species	KMeans
1	5.1	3.5	1.4	0.2	setosa	3
2	4.9	3.0	1.4	0.2	setosa	3
3	4.7	3.2	1.3	0.2	setosa	3
4	4.6	3.1	1.5	0.2	setosa	3
5	5.0	3.6	1.4	0.2	setosa	3
6	5.4	3.9	1.7	0.4	setosa	3
7	4.6	3.4	1.4	0.3	setosa	3
8	5.0	3.4	1.5	0.2	setosa	3
9	4.4	2.9	1.4	0.2	setosa	3
10	4.9	3.1	1.5	0.1	setosa	3
11	5.4	3.7	1.5	0.2	setosa	3
12	4.8	3.4	1.6	0.2	setosa	3
13	4.8	3.0	1.4	0.1	setosa	3
14	4.3	3.0	1.1	0.1	setosa	3
15	5.8	4.0	1.2	0.2	setosa	3
16	5.7	4.4	1.5	0.4	setosa	3
17	5.4	3.9	1.3	0.4	setosa	3
18	5.1	3.5	1.4	0.3	setosa	3
19	5.7	3.8	1.7	0.3	setosa	3
20	5.1	3.8	1.5	0.3	setosa	3
50	5.0	3.3	1.4	0.2	setosa	3
51	7.0	3.2	4.7	1.4	versicolor	1
52	6.4	3.2	4.5	1.5	versicolor	1
53	6.9	3.1	4.9	1.5	versicolor	2
54	5.5	2.3	4.0	1.3	versicolor	1
55	6.5	2.8	4.6	1.5	versicolor	1
56	5.7	2.8	4.5	1.3	versicolor	1
57	6.3	3.3	4.7	1.6	versicolor	1

100	5.7	2.8	4.1	1.3	versicolor	1
101	6.3	3.3	6.0	2.5	virginica	2
102	5.8	2.7	5.1	1.9	virginica	1
103	7.1	3.0	5.9	2.1	virginica	2
104	6.3	2.9	5.6	1.8	virginica	2
105	6.5	3.0	5.8	2.2	virginica	2
106	7.6	3.0	6.6	2.1	virginica	2
107	4.9	2.5	4.5	1.7	virginica	1
108	7.3	2.9	6.3	1.8	virginica	2
109	6.7	2.5	5.8	1.8	virginica	2
110	7.2	3.6	6.1	2.5	virginica	2
111	6.5	3.2	5.1	2.0	virginica	2
112	6.4	2.7	5.3	1.9	virginica	2
113	6.8	3.0	5.5	2.1	virginica	2
114	5.7	2.5	5.0	2.0	virginica	1
115	5.8	2.8	5.1	2.4	virginica	1

3: setosa
2: virginica
1: versicolor

ほぼ、このような分類で
クラスタリングされている

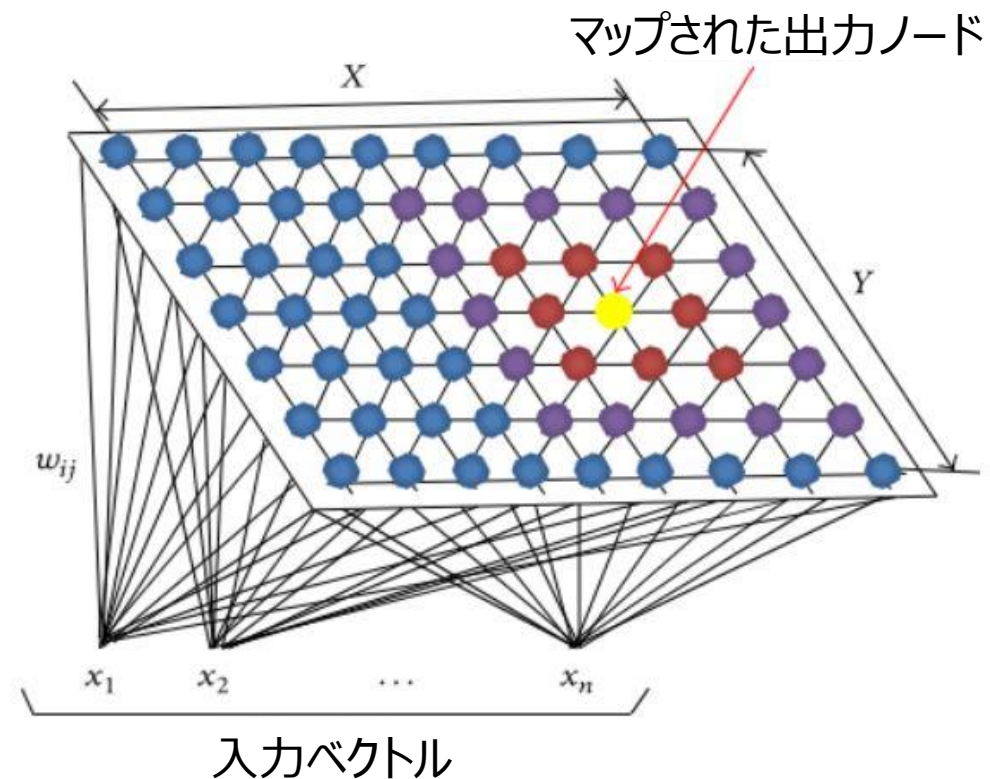
自己組織化マップとは？

自己組織化マップ° (Self-organizing Map) とは、ニューラルネットワークのアルゴリズムの一つで、多次元のデータを2次元平面状へ非線形投射する手法である。

p次元の入力ベクトル \mathbf{X} が、2次元(x-y平面)上にあるノードのどこかにマッピングされる

$$\|\mathbf{x}_j - \mathbf{m}_c\| = \min_i \{\|\mathbf{x}_j - \mathbf{m}_i\|\}$$

これを満たすような重み(\mathbf{m}_i)ベクトルを算出する

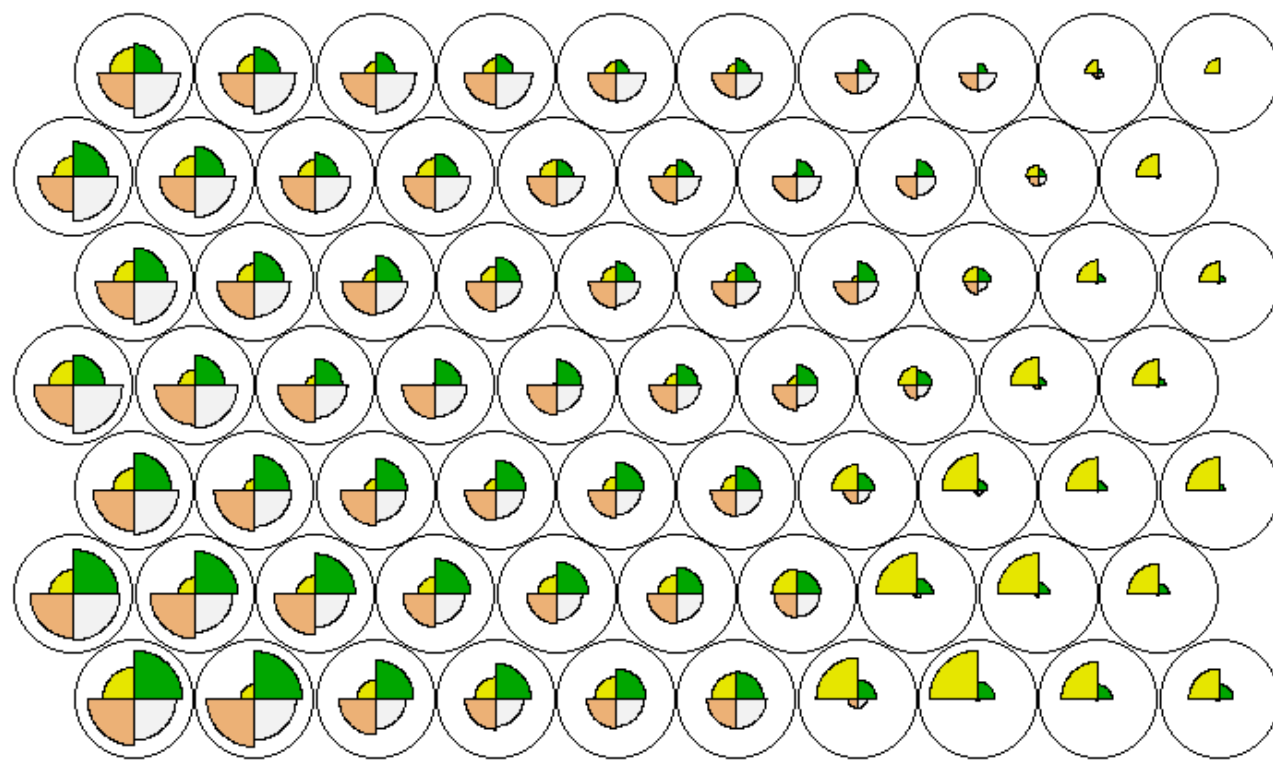


SOMによる自己組織化マップ例

irisデータを用いてアヤメを分類
kohonenが未インストールの場合は、まず
Install.packages("kohonen")を実行

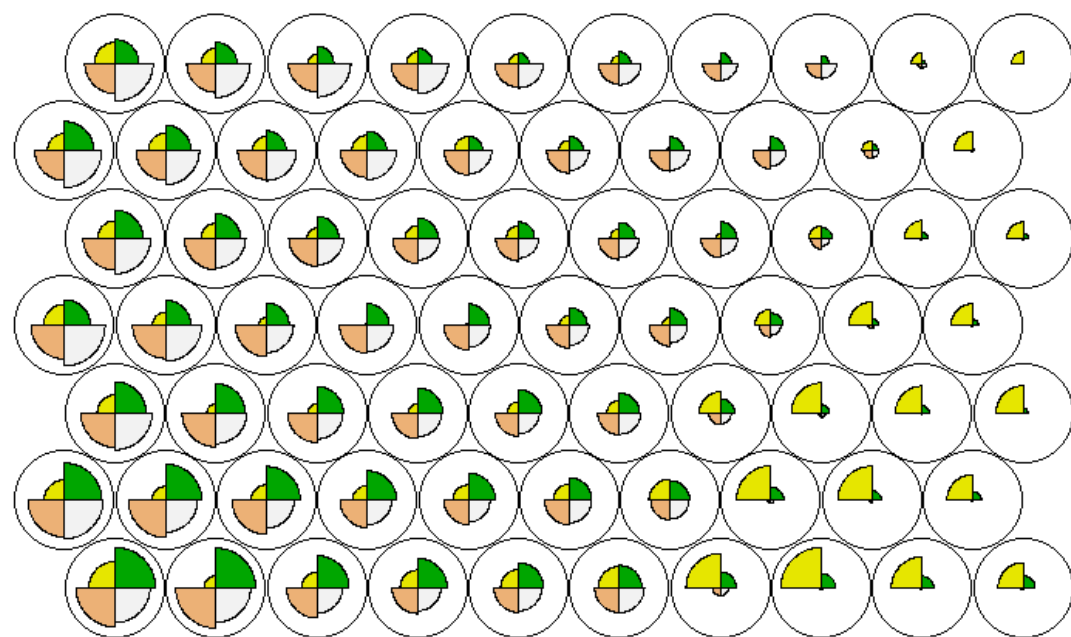
```
> library(kohonen)
> set.seed(10)
> gr <- somgrid(topo="hexagonal", xdim=10,ydim=7)
> iris.som <- som(as.matrix(iris[,1:4]),gr,rlen=200)
> plot(iris.som)
```

irisデータの1:4データを用いて、**SOM**により
10 x 7グリッドに各サンプルが振り分けられる



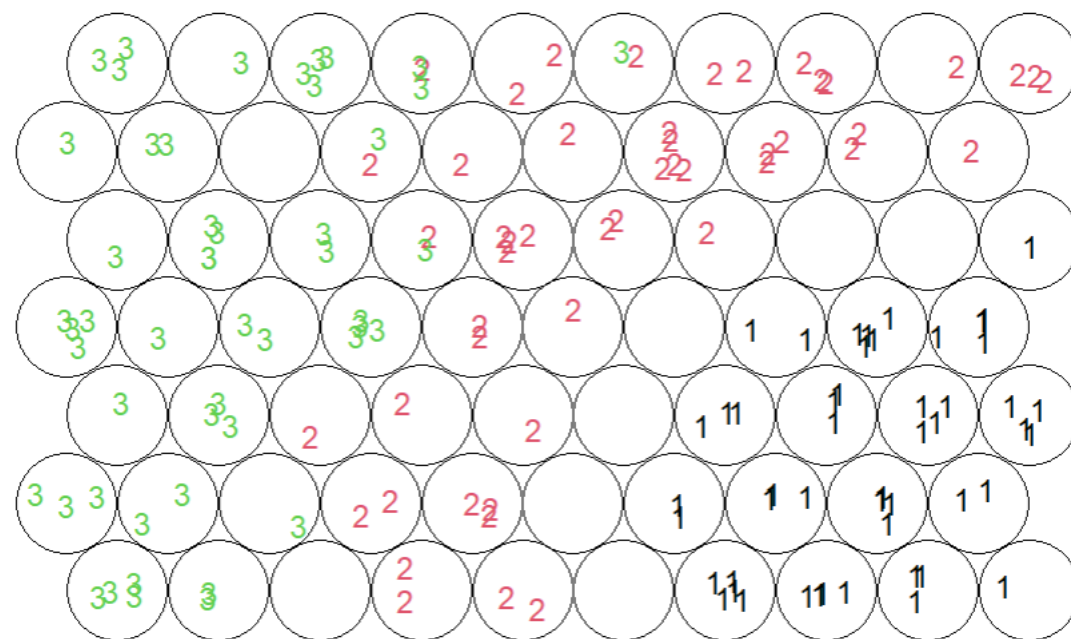
■ Sepal.Length	■ Petal.Length
■ Sepal.Width	■ Petal.Width

自己組織化マップとラベルの対応づけ



■ Sepal.Length ■ Petal.Length
■ Sepal.Width ■ Petal.Width

```
> lab.cod<- as.numeric(iris[,5])  
> plot(iris.som, type="mapping", labels=lab.cod, col=lab.cod)
```



1: setosa, 2: versicolor, 3: virginica

SPAM 1 データ

```
> spam1 <- read.csv("spam1.csv")
```

大文字割合 総大文字数

	remove	free	email	you	charExclamation	capitalAve	capitalTotal	type
1	0.00	0.32	1.29	1.93	0.778	3.756	278	spam
2	0.21	0.14	0.28	3.47	0.372	5.114	1028	spam
3	0.19	0.06	1.03	1.36	0.276	9.821	2259	spam
4	0.31	0.31	0.00	3.18	0.137	3.537	191	spam
5	0.31	0.31	0.00	3.18	0.135	3.537	191	spam
6	0.00	0.00	0.00	0.00	0.000	3.000	54	spam
7	0.00	0.96	0.32	3.85	0.164	1.671	112	spam
8	0.00	0.00	0.00	0.00	0.000	2.450	49	spam
9	0.30	0.00	0.15	1.23	0.181	9.744	1257	spam
10	0.38	0.00	0.12	1.67	0.244	1.729	749	spam
11	0.96	0.00	0.96	3.84	0.462	1.312	21	spam
12	0.25	0.00	0.00	1.16	0.663	1.243	184	spam
13	0.00	0.34	1.39	2.09	0.786	3.728	261	spam
14	0.90	0.00	0.00	2.72	0.000	2.083	25	spam
15	0.00	5.35	0.00	3.21	0.357	1.971	205	spam
16	0.42	1.27	0.00	1.70	0.572	5.659	249	spam
17	0.00	0.00	0.00	1.88	0.428	4.652	107	spam
18	0.00	0.00	0.00	0.00	1.975	35.461	461	spam
19	0.18	0.00	0.37	3.15	0.455	1.320	70	spam
20	0.00	0.63	3.18	2.22	0.055	3.509	186	spam

remove

→ 削除されます…

free

→ いまなら無料…

email

→ メールして…

you

→ あなただけに…

! 数

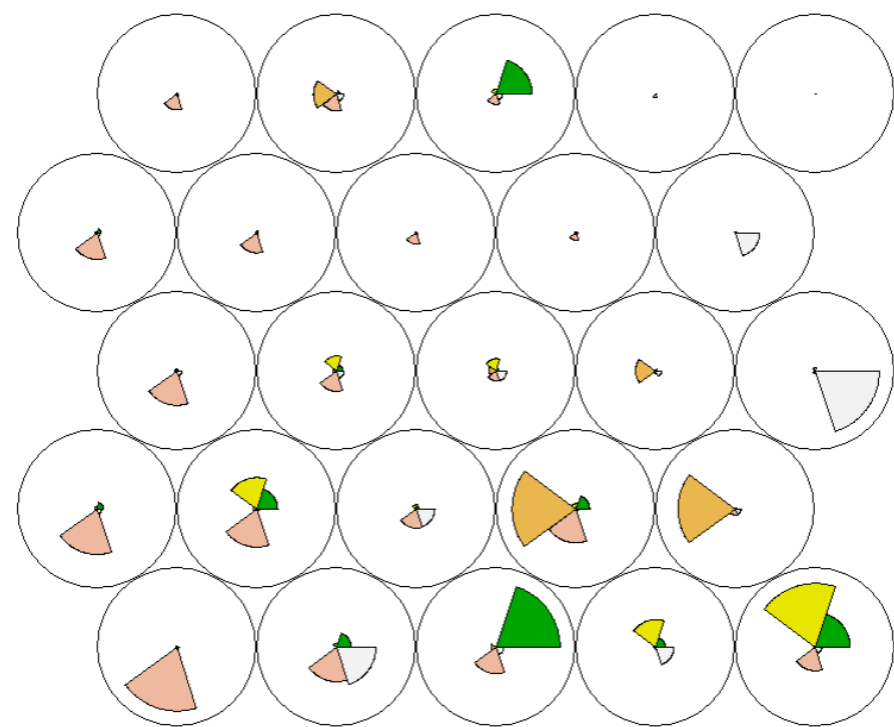
→ 大変お得的な…

SPAMデータを用いた自己組織化マップ

```
> gr <- somgrid(topo="hexagonal", xdim=5,ydim=5)  
> spam1.som <- som(as.matrix(spam1[,1:5]),gr,rlen=200)  
> plot(spam1.som)
```

5 x 5 グリッドに5つの特徴量で**自己組織化マップ**を出力

→ 丸の中の扇の大きさは各特徴量の大きさを表す！



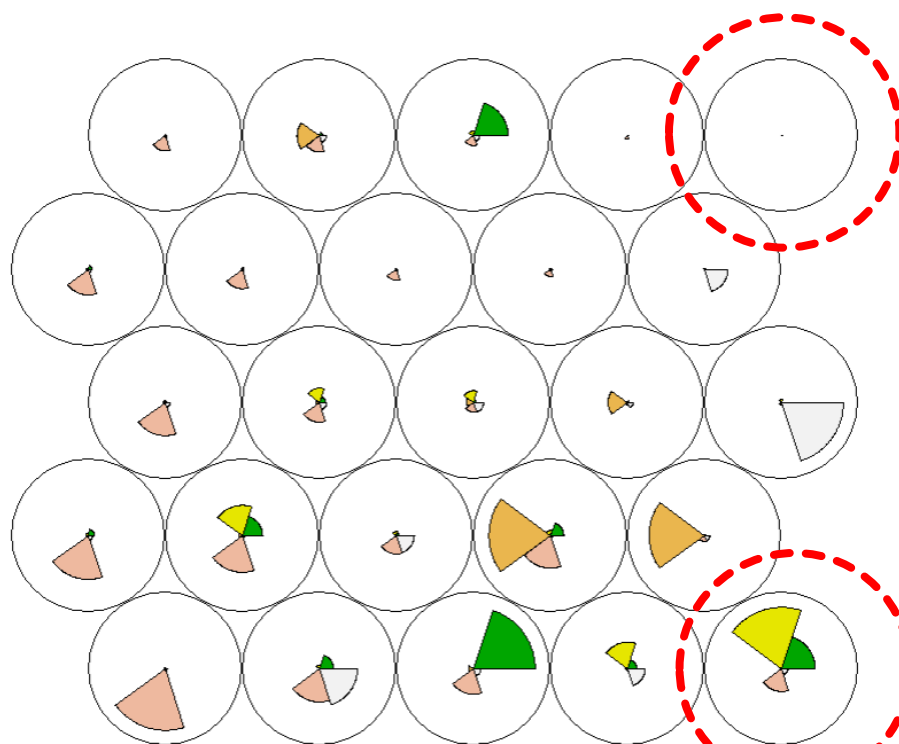
■ remove
■ free

■ email
■ you

■ charExclamation

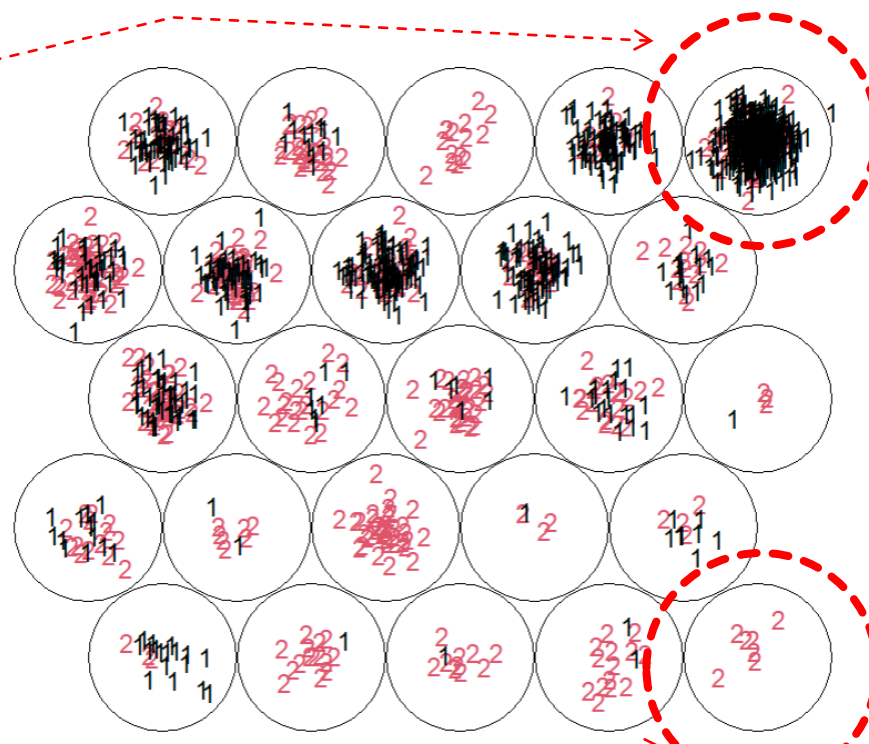
自己組織化マップとラベルの対応づけ

```
> lab.cod1 <- as.numeric(as.factor(spam1[,8]))  
> plot(spam1.som, type="mapping", labels=lab.cod1, col=lab.cod1)
```



■ remove	■ email	□ charExclamation
■ free	■ you	

正常メールはシンプル！



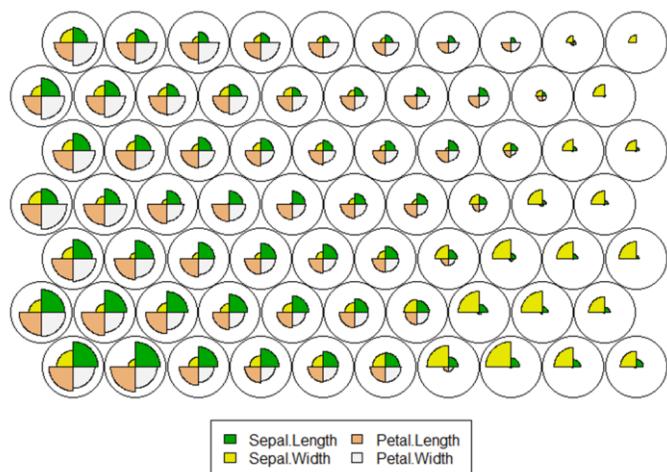
迷惑メールは甘い話！

1: Ham, 2: Spam

課題：SPAM 1 データを用いた拡張自己組織化マップ

SPAM1には、演習で採用した特徴量(remove, free, you, email, ! 数)以外にも、**大文字割合**、**総大文字数**のデータも記録されています。特徴量を7つまで拡張し、グリッドを10 x 7に設定して自己組織化マップを作成して、スパムメールの特徴について考察下さい。

irisデータに対する10x7グリッドの例



	remove	free	email	you	charExclamation	大文字割合	総大文字数	type
						capitalAve	capitalTotal	
1	0.00	0.32	1.29	1.93	0.778	3.756	278	spam
2	0.21	0.14	0.28	3.47	0.372	5.114	1028	spam
3	0.19	0.06	1.03	1.36	0.276	9.821	2259	spam
4	0.31	0.31	0.00	3.18	0.137	3.537	191	spam
5	0.31	0.31	0.00	3.18	0.135	3.537	191	spam
6	0.00	0.00	0.00	0.00	0.000	3.000	54	spam
7	0.00	0.96	0.32	3.85	0.164	1.671	112	spam
8	0.00	0.00	0.00	0.00	0.000	2.450	49	spam
9	0.30	0.00	0.15	1.23	0.181	9.744	1257	spam
10	0.38	0.00	0.12	1.67	0.244	1.729	749	spam
11	0.96	0.00	0.96	3.84	0.462	1.312	21	spam
12	0.25	0.00	0.00	1.16	0.663	1.243	184	spam
13	0.00	0.34	1.39	2.09	0.786	3.728	261	spam
14	0.90	0.00	0.00	2.72	0.000	2.083	25	spam
15	0.00	5.35	0.00	3.21	0.357	1.971	205	spam
16	0.42	1.27	0.00	1.70	0.572	5.659	249	spam
17	0.00	0.00	0.00	1.88	0.428	4.652	107	spam
18	0.00	0.00	0.00	0.00	1.975	35.461	461	spam
19	0.18	0.00	0.37	3.15	0.455	1.320	70	spam
20	0.00	0.63	3.18	2.22	0.059	3.509	186	spam



データマイニングを楽しもう！