

データマイニング入門

Rで学ぶ最新データ解析 東京図書 2008年

74ページよりの2.5.2項および2.5.3項の差し替え原稿

2014年11月19日

早稲田大学 豊田秀樹 編著

2.5.2 鉛筆の数え方データの分析

大石 鉛筆の数え方を学習するモデルでは、どのようなネットワークが設定されていたのかしら？

浅野 入力層のユニットに1から10の「本数」を、出力層のユニットに「ボン」「ホン」「ボン」という3通りの教え方を設定して、その間に中間層を1層設定したモデルになっていました。このモデルでは、中間層の複数のユニットに対して、ロジスティック関数（シグモイド変換）ではなく、ガウシアン関数（ガウシアン変換）を設定して学習を行っているんです。

大石 S字型じゃなくて、釣鐘型の関数を利用したわね。でも、ここではS字型のロジスティック関数を使うことにしましょう。

浅野 どうしてですか。図2.15のような形状の関数は、S字型ではなく釣鐘型の関数の和のほうが近似が容易だったのではありませんか。

大石 そのとおりよ。この本が公刊された2008年には“neural”というパッケージがあつて、ガウシアン関数を使ったり、多層の隠れユニットを構成したりできたわ。今でも(2014年11月現在)、“neural”というパッケージは使えます。でも最新のR環境(R-3.1.2)ではパッケージ“neural”は使えないの。ここではその代わりに“neuralnet”というパッケージ(Ver.1.32)を紹介するわ。でも、残念ながら“neuralnet”では釣鐘型の関数は使えません。

浅野 S字型のロジスティック関数で大丈夫なんですか？

```
library(neuralnet)
鉛筆データ<-read.csv("鉛筆.csv",header=T)
鉛筆データ$偶奇<-鉛筆データ$本数%%2
```

大石 予測変数を増やせば大丈夫。まずlibrary()でパッケージを使えるようにして、read.csv()でデータを読み込みます。割算のあまりを求める演算子%%を利用して、鉛筆の本数が偶数か奇数かを与える変数「偶奇」を作ります。

浅野 なるほど！「本数」だけではなく、それが偶数か奇数かということも利用して、数え方をネットに覚えさせるんですね。

```
set.seed(4)
print(net.鉛筆 <- neuralnet(ボン+ホン+ボン~本数+偶奇, 鉛筆データ, hidden=3))
```

大石 そのとおり。まずset.seed()で分析結果を再現できるようにします。パッケージ“neuralnet”のメインの関数は同名のneuralnet()よ。この関数でネット学習をするの。第1引数はモデル式の指定で、チルダ~の左辺が基準変数(教師情報)、右辺が予測変数ね。第2引数はデータフレーム。第3引数では、隠れ層のユニット数を指定しているわ。

浅野 大丈夫かなあ。あ、もう結果がでた。えーとErrorが0.0053と書いてある。

大石 教師情報は 0-1 変数だから、Error が 0.0053 なら、学習は完璧ということね。

```
plot(net. 鉛筆, rep="best")
plot(net. 鉛筆, show.weights=F)
```

浅野 plot() でネットワークポロジリーが確認できるのですね。おや、引数に show.weights=F を指定すると、重みを省略できるようだぞ。

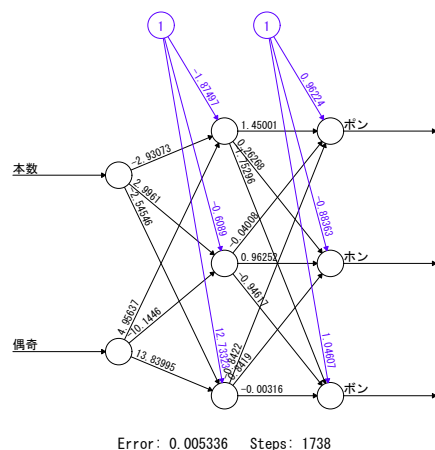


図 2.17 鉛筆の数え方を学習したネット

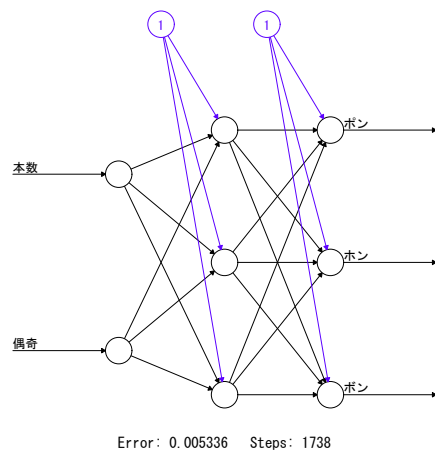


図 2.18 鉛筆の数え方を学習したネット (重みなし)

```
confidence.interval(net. 鉛筆)
print(pre. 鉛筆<-prediction(net. 鉛筆))
print(err. 鉛筆<-round(pre. 鉛筆 [[2]]-pre. 鉛筆 [[1]],2))
print(cur. 鉛筆<-compute(net. 鉛筆, cbind(c(1:10),c(1:10)%2))$net.result)
```

大石 モデルが識別されていて、隠れ層が 1 層の場合は、関数 confidence.interval() で重みの信頼区間を求めることもできるし、予測値は関数 prediction() のリストの 1 番目で得られ

るわ。リストの2番目には教師情報があるから、差を求めれば誤差が計算できるの。

浅野 学習済みのニューラルネットを使って、新たなデータの予測ができますか？

大石 できるわ。関数 `compute()$net.result` を使い、第2引数に新たなデータを置けばOKよ。ここでは同じデータを置いて、その結果を使って学習された曲線を図 2.1.9 に示したわ。

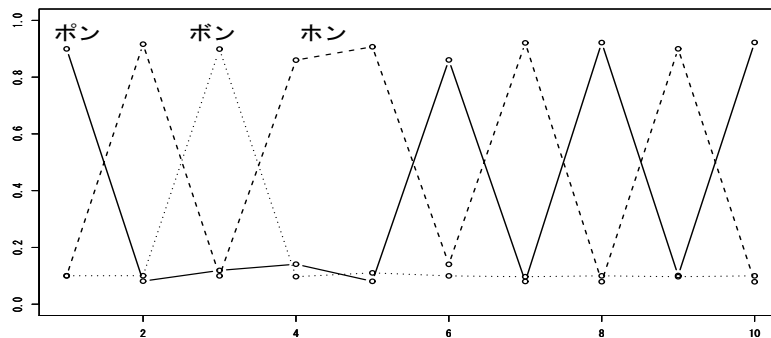


図 2.19 学習結果の描画

浅野 たとえば1, 6, 8, 10の時に「ポン」と数える予測値が高くなっているから、学習がきちんと行われているようですね。（ぐ〜っ…、とお腹がなる。）すみません、真剣に勉強したらなんかお腹が空いたみたいで。

大石 じゃあ、持ってきてくれたケーキの箱をそろそろ開けましょうか。

2.5.3 ケーキの購入意思データの分析

大石 まあ、おいしそう！たくさんあって目移りしちゃうわ。浅野くんはどれがいいかしら？

浅野 ここは先輩から先に選んでください。

大石 遠慮なんかすることないのに。じゃあ、「せーの！」で食べたいケーキを指差しましょう。

大石・浅野 せーの！

浅野 分れましたね。俺と先輩の頭の中では、食べたいケーキに関して異なるネットワークが張られてるみたいですね。

大石 そうね、ものの見方や感じ方は人によってさまざまだから、どのケーキを選ぶかはそれを判断する人によって異なるわね。ニューラルネットは脳内の神経構造を模したものだから、当然、主観的な思考や判断に関するルールも学習できるわよ。

そういえば、ケーキの購入意思に関するデータが丁度あるんだけど、このデータを用いてニューラルネット分析してみない？

浅野 是非やりましょう！今度は、中間層を2層以上設定してネットワークモデルの学習を行ってみたいんです。

大石 いいわよ。まず、分析の前にデータの説明をしておくわね。このデータは、私の友達の浮橋うきはしさんが、「高級感」、「甘さ」、「新奇性」、「濃厚さ」、「食感（やわらかさ）」、「華やかさ」、「値段」という7つの観点から35種類のケーキに対する「購入意思」を評価したものなの（図 2.20）。

cake の種類, 高級感, 甘さ, 新奇性, 濃厚さ, 食感, 華やかさ, 値段, 購入意思

```

cake01,2,3,1,3,4,5,280,1
cake02,4,2,4,2,5,4,280,1
cake03,4,4,1,4,2,4,260,0
cake04,3,5,1,5,4,2,280,0
cake05,4,4,1,4,2,3,300,0
cake06,3,4,2,5,2,1,300,0
cake07,3,3,3,3,3,3,100,0
cake08,2,4,1,3,4,1,280,0
cake09,2,4,2,4,4,1,280,0
cake10,5,4,5,4,3,5,300,1
cake11,2,5,4,4,2,3,300,0
cake12,4,3,4,4,3,4,320,1
cake13,4,4,1,4,3,3,280,0
cake14,5,4,3,4,3,5,360,1
cake15,3,4,1,4,2,3,400,0
cake16,4,3,1,4,2,3,400,0
cake17,2,4,1,4,2,2,400,0
cake18,5,3,2,3,2,5,380,1
cake19,5,3,4,3,3,4,320,0
cake20,5,3,5,3,3,3,320,1
cake21,1,2,2,2,2,2,350,0
cake22,2,2,1,1,2,1,160,0
cake23,5,4,5,3,3,5,330,1
cake24,5,4,5,3,3,4,330,1
cake25,3,5,1,5,5,1,110,0
cake26,4,5,2,5,5,1,120,0
cake27,1,4,4,4,3,4,110,1
cake28,1,4,4,4,3,4,110,1
cake29,1,5,4,4,3,4,110,0
cake30,2,3,1,3,2,3,120,0
cake31,2,3,3,3,2,3,120,0
cake32,4,2,3,3,4,2,300,0
cake33,5,1,4,2,2,5,360,1
cake34,5,4,5,4,3,5,320,1
cake35,3,4,3,3,3,2,300,0

```

図 2.20 浮橋さんのケーキに対する評価データ

大石 値段以外の 6 つの観点は、まったくないと思う場合を 1 点、とてもあると思う場合を 5 点とする 5 件法によって評価しているわよ。「購入意思」に関しては、買いたいと思うならば 1、買いたくないと思うならば 0 と評定しているわ。ここでは、観点に関する 7 つの変数を予測変数に、「購入意思」を基準変数に設定しましょう。

浅野 本当に浮橋さんの頭の中のネットワークを学習できるのかなあ。

大石 楽しみね。じゃあ、早速 MLP モデルの分析を実行していきましょう。

浅野 始めに、パッケージとデータの読み込みを行うんですね。

```

library(neuralnet)
ケーキデータ<-read.csv("ケーキ.csv",header=T)

```

大石 よしよし。次は、7 つの予測変数をそれぞれ 0 から 1 の範囲にスケーリングしましょう。

```
for (i in 2:8){
  ケーキデータ[,i]<-(ケーキデータ[,i]-min(ケーキデータ[,i]))/
  (max(ケーキデータ[,i])-min(ケーキデータ[,i]))
}
ケーキデータ
```

浅野 これでいいでしょうか。

大石 よくできました。続いて、関数 `neuralnet()` で入力層・中間層・出力層のユニットを指定しましょう。

```
set.seed(5)
print(net. ケーキ
      <- neuralnet(購入意思~高級感+甘さ+新奇性+濃厚さ+食感+華やかさ+値段,
                    ケーキデータ, hidden=c(2,1)))
```

浅野 チルダ(~)の左辺が出力層(教師情報・基準変数)、右辺が入力層(予測変数)でしたね。

大石 ええ。ここでは中間層を2層仮定して、一層めのユニット数を2つ、二層めのユニット数を1つ設定してみましょう。引数 `hidden=` に、ベクトルで指定します。

浅野 えーと `Error` が 0.00498 と書いてある。教師情報は 0-1 変数だから、`Error` が 0.00498 なら、学習は完璧と考えていいということでしたね。

大石 鉛筆の学習と同じように、重みのある図とない図を用いて、ネットワークトポロジーを確認しておきましょう。

```
plot(net. ケーキ, rep="best")
plot(net. ケーキ, show.weights=F)
```

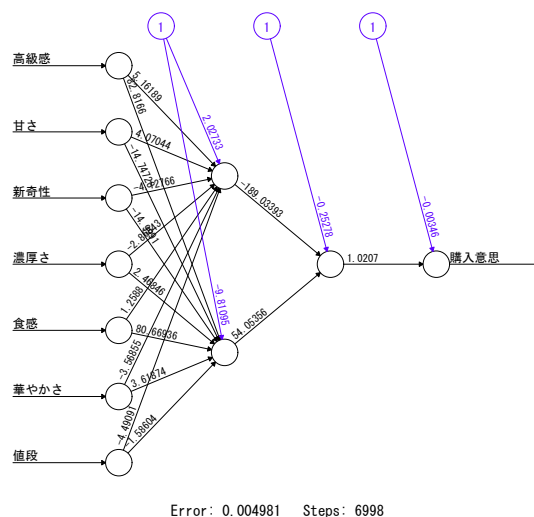


図 2.21 「購入意思」を学習したネット

浅野 ああ、確かに隠れ層が2層になっていますね。でもこのネットは、本当に浮橋さんの好み

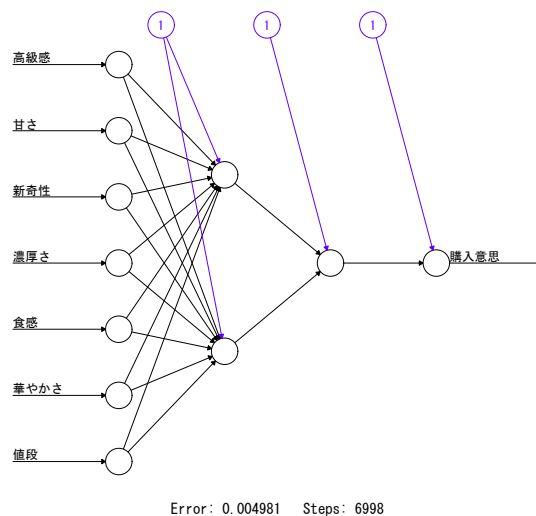


図 2.22 「購入意思」を学習したネット (重みなし)

を学習したのでしょうか？

大石 じゃあ確認してみましょう。

```
cur. ケーキ<-compute(net. ケーキ, ケーキデータ [,2:8])$net.result
result <- cbind(ケーキデータ [,9],round(cur. ケーキ,2))
dimnames(result)<-list(ケーキデータ [,1],c("購入意思","予測値"))
t(result)
```

浅野 えーと compute() で教師情報を予測して、cbind() で教師情報とその予測値をまとめて、行名・列名をつけてっと。

大石 整形が上手になったわね。

	cake01	cake02	cake03	cake04	cake05	cake06	cake07	cake08
購入意思	1.00	1.00	0	0	0	0	0	0
予測値	1.02	0.98	0	0	0	0	0	0
	cake09	cake10	cake11	cake12	cake13	cake14	cake15	cake16
購入意思	0	1.00	0	1.00	0	1.00	0	0
予測値	0	1.02	0	1.02	0	0.99	0	0
	cake17	cake18	cake19	cake20	cake21	cake22	cake23	cake24
購入意思	0	1.00	0.00	1.00	0	0	1.00	1.00
予測値	0	0.97	0.07	0.96	0	0	1.02	0.99
	cake25	cake26	cake27	cake28	cake29	cake30	cake31	cake32
購入意思	0	0	1	1	0	0	0	0
予測値	0	0	1	1	0	0	0	0
	cake33	cake34	cake35					
購入意思	1.00	1.00	0					
予測値	1.02	1.02	0					

浅野 すごい！35 個のケーキの「購入意思」を完璧に学習している。

大石 まあね。だけどこのネットが浮橋さんの好みを本当に学習したかどうかは、実際にはまだ分からないのよ。ネットの学習に利用していないケーキのデータを入力したとき、それでも浮橋さんの判断と一致することが大切なんだからね。

浅野 はい、交差妥当性の確認（交差検証）でしたね。わかりました。… ふ〜っ。これで、R を使ったニューラルネット分析の方法を 3 つ覚えたぞ！いつの間にか日が暮れてしまいましたね。

大石 まあ、本当。浅野くんはヤル気があって教え甲斐があるわね。

浅野 実際にネットワークを構築できるのって楽しくて、つい夢中になってしまいました。

大石 もし興味を持ったなら、上記のプログラムの指定を書き換えたり、身近にある他のデータを用いたりして、さらに分析を行ってみてね。

浅野 是非やってみます。ありがとうございました。次は決定木の理論を勉強しようと思っているので、今度は R を使った決定木分析の講義をしてもらえたら嬉しいんですけど。

大石 いいわよ、また遠慮なくいらっしゃい。これで浅野くんもデータマイナーとしての一步を踏み出したわね。

浅野 はい、頑張ります！ところで、おやつを食べて間もないですけど、これから夕飯にいきませんか？

大石 いいわよ。甘いものを食べたから、代わりにラーメンでも食べたいわね。

— 二人、研究室を去る —