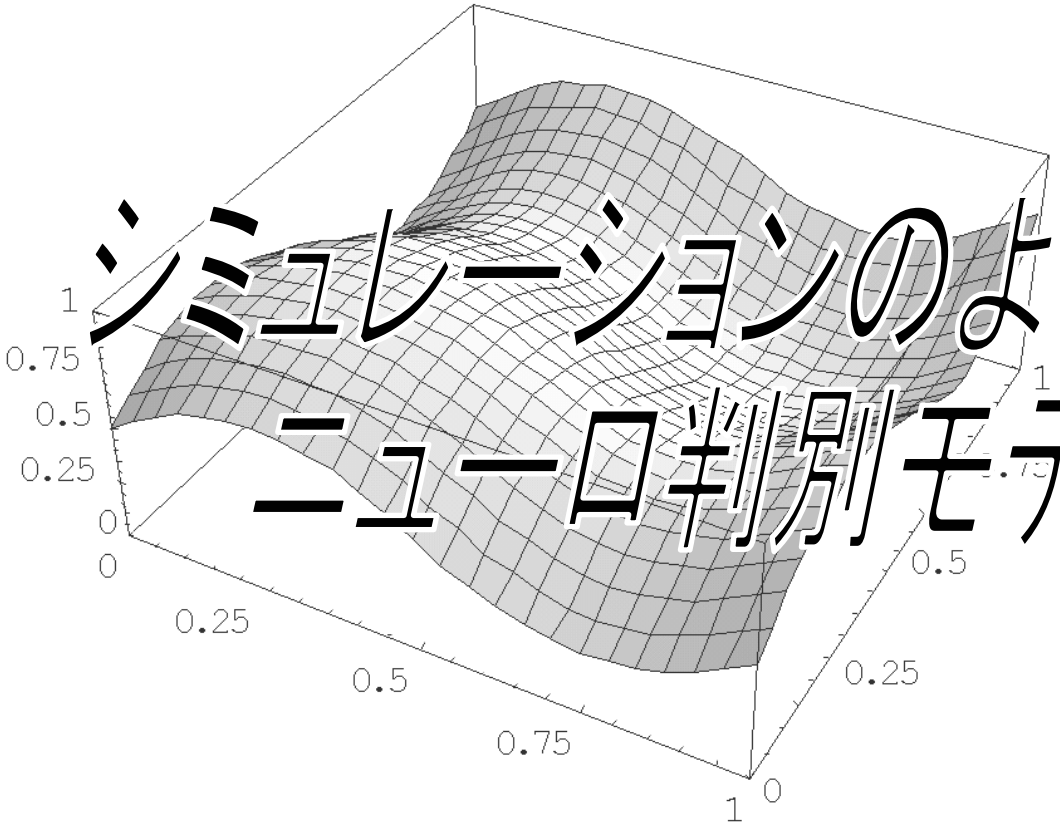
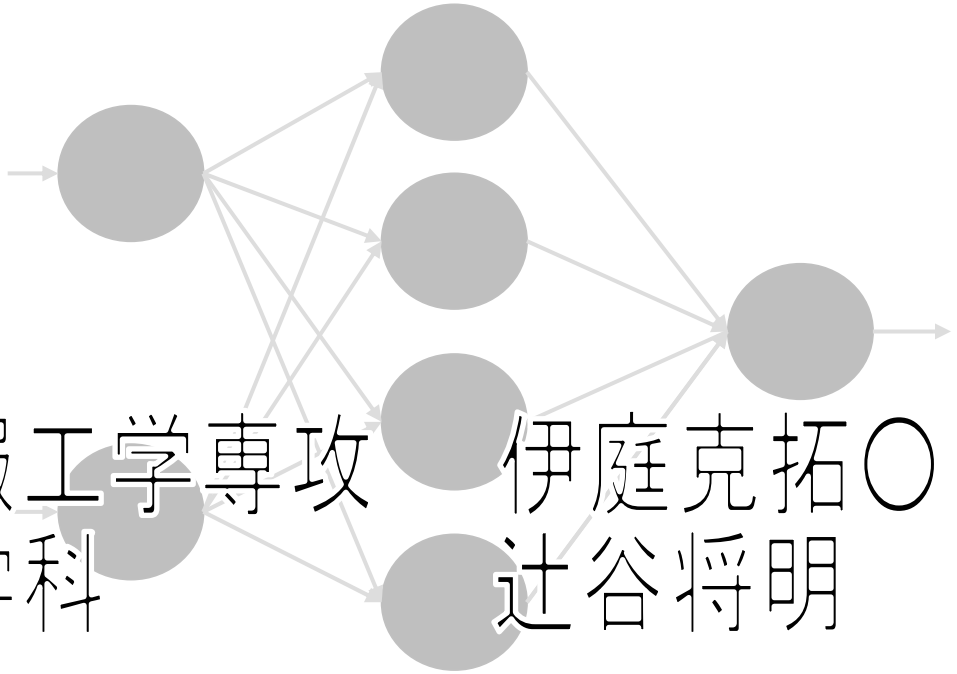


シミュレーションのよる ニューロ判別モデルの性能評価



大阪電気通信大学
大学院工学研究科情報工学専攻　伊庭克拓○
総合情報学部情報工学科　辻谷将明

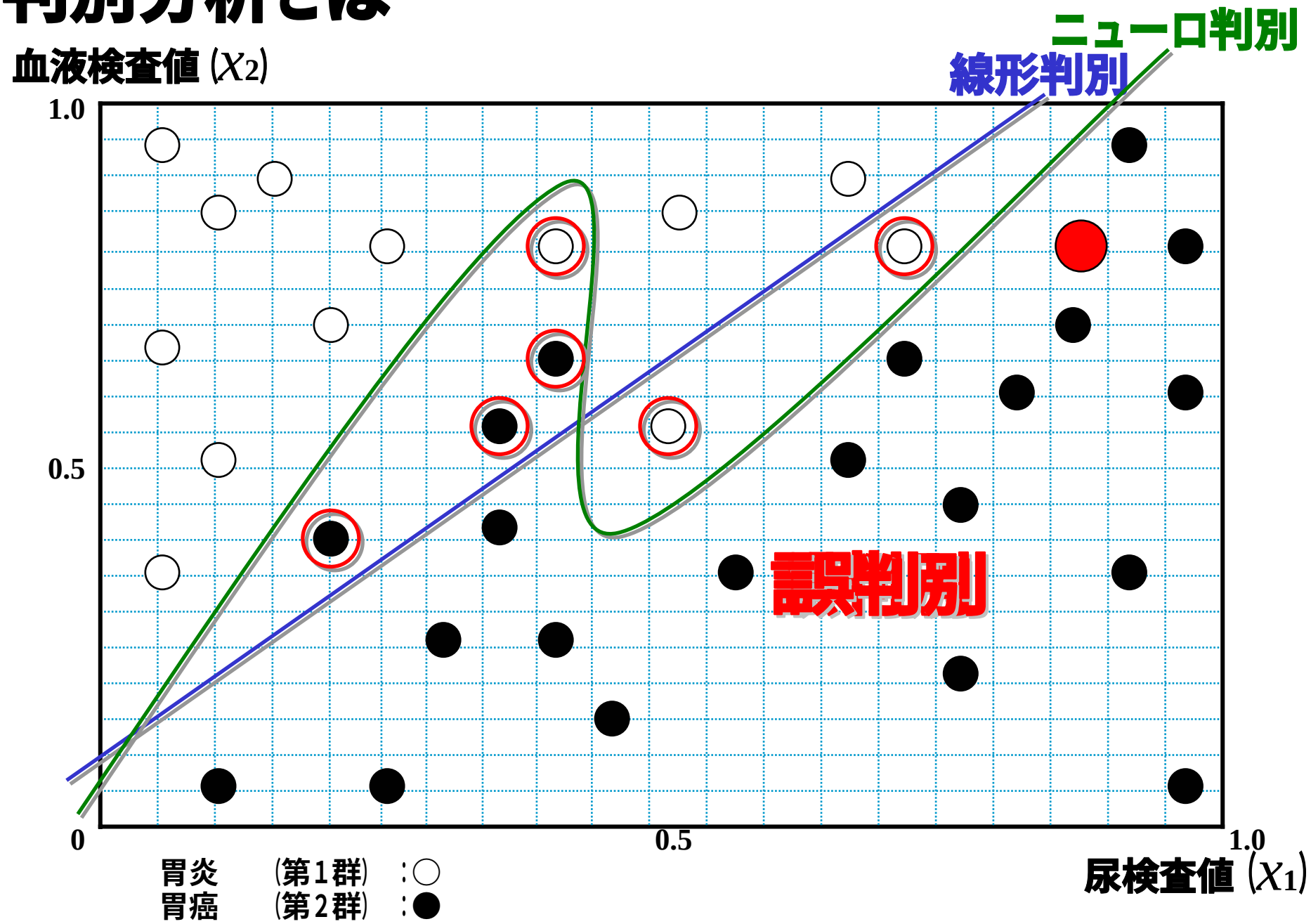
1. 研究背景

- 近年，計算機システムの性能の向上により，
  高度な非線形問題へのアプローチ
  を，従来の解析的手法と計算機システムを組み合わせた手法によって行なう研究が活発化.

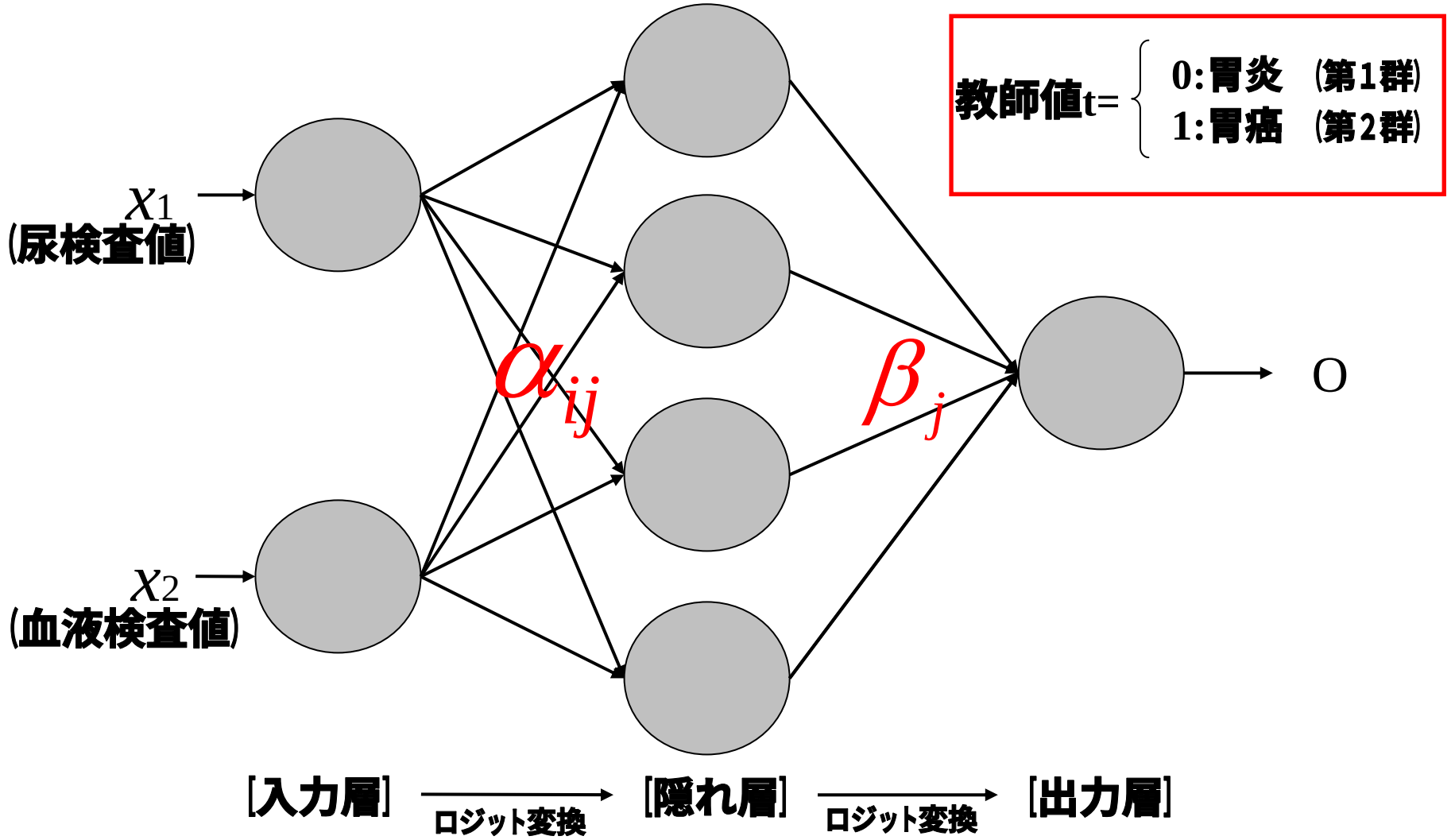
- **ニューラルネットワーク**
  とは，人間の脳の神経系を模倣した計算機システムで，入力に対する理想的な出力である教師値によって学習を行い，任意の関数を近似することができる.

ニューラルネットに基づく**非線形**判別分析モデルの
シミュレーションデータによる研究.

判別分析とは



ニューロ判別モデル（2群判別）



モデル構築

［入力層］ —ロジット変換→ ［隠れ層］ —ロジット変換→ ［出力層］

- ［入力層］ —ロジット変換→ ［隠れ層］

$$u_j = \sum_{i=0}^{I} \alpha_{ij} x_i \Rightarrow \quad y_j = \frac{1}{1+\exp(-u_j)}$$

- ［隠れ層］ —ロジット変換→ ［出力層］

$$v = \sum_{j=0}^{J} \beta_j y_j \Rightarrow \quad o = \frac{1}{1+\exp(-v)}$$

［出力］

ニューラルネットによる関数の近似

ニューロ判別モデル（2群判別）

- 出力と教師値の誤差より未知パラメータを推定.
- 隠れユニット数を増やすことにより任意の関数を近似.

$$o = \frac{1}{1 + \exp\left[-\sum_{j=0}^{J} \left\{ \frac{\beta_j}{1 + \exp\left(-\sum_{i=0}^{I} \alpha_{ij} x_i \right)} \right\} \right]}$$

シミュレーションの必要性

患者番号	生後日数 (x_1)	手術した脊柱の個数 (x_2)	群 (t)
1	75	3	0
2	158	3	0
3	128	4	1
・	・	・	・
・	・	・	・
81	36	4	**0**

過去の結果
専門家の経験

脊柱後湾症

脊柱後湾症は脊柱の生理的湾曲のうち,胸椎部の後方凸湾が,より後方に変形した症状で,頭部,および上下肢を支えるのに支障をきたす.

例えば,群がこのような関数で決まっていても,

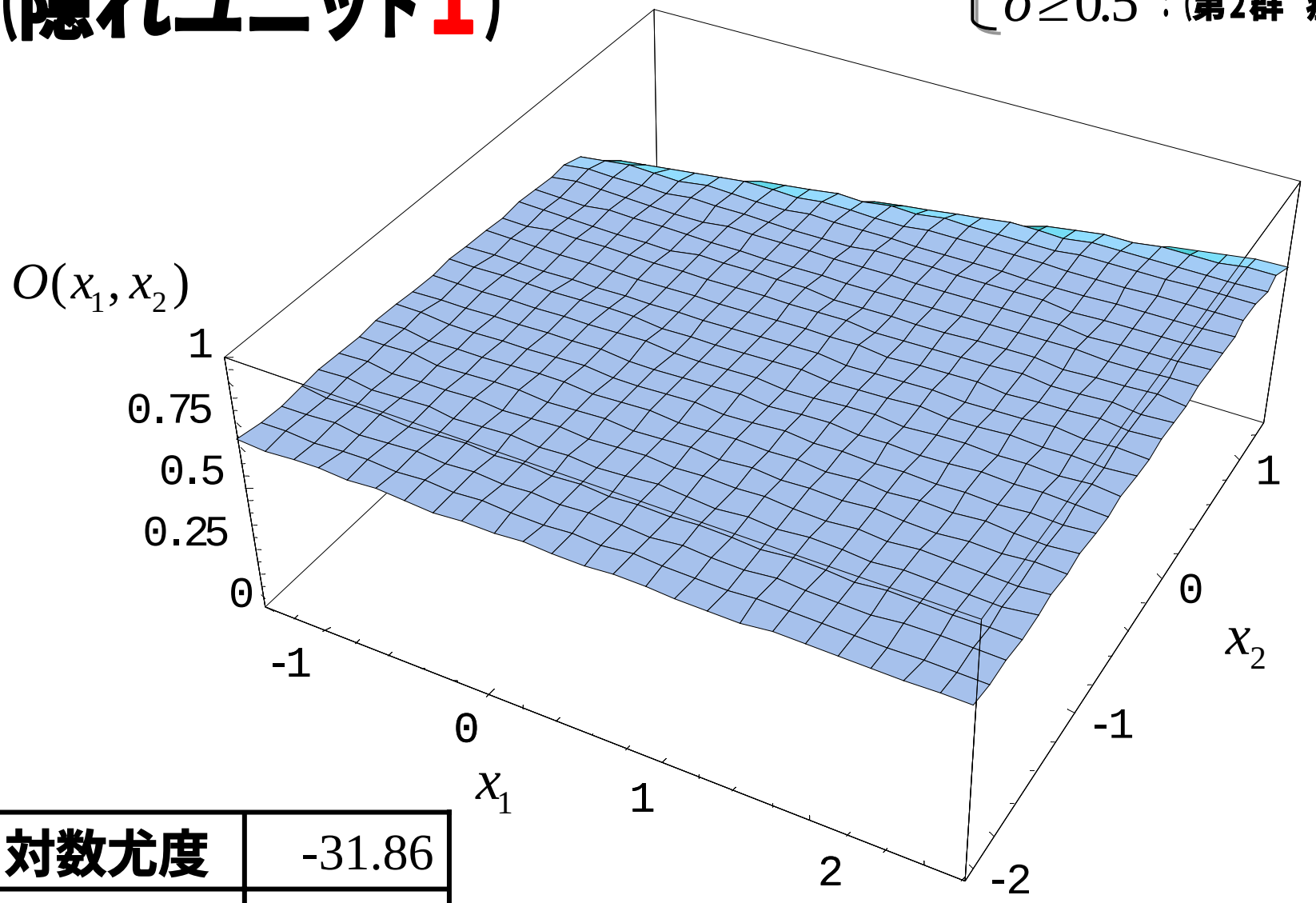
$$P_r = f(x_1, x_2)$$

$$\text{群 (教師値 t)} = \begin{cases} t=0 \ (\text{第1群} \quad \text{症状なし}) & P_r < 0.5 \\ t=1 \ (\text{第2群} \quad \text{症状あり}) & P_r \geq 0.5 \end{cases}$$

実際のデータでは、この母集団の関数は未知

ニューロ判別モデルの出力（隠れユニット1）

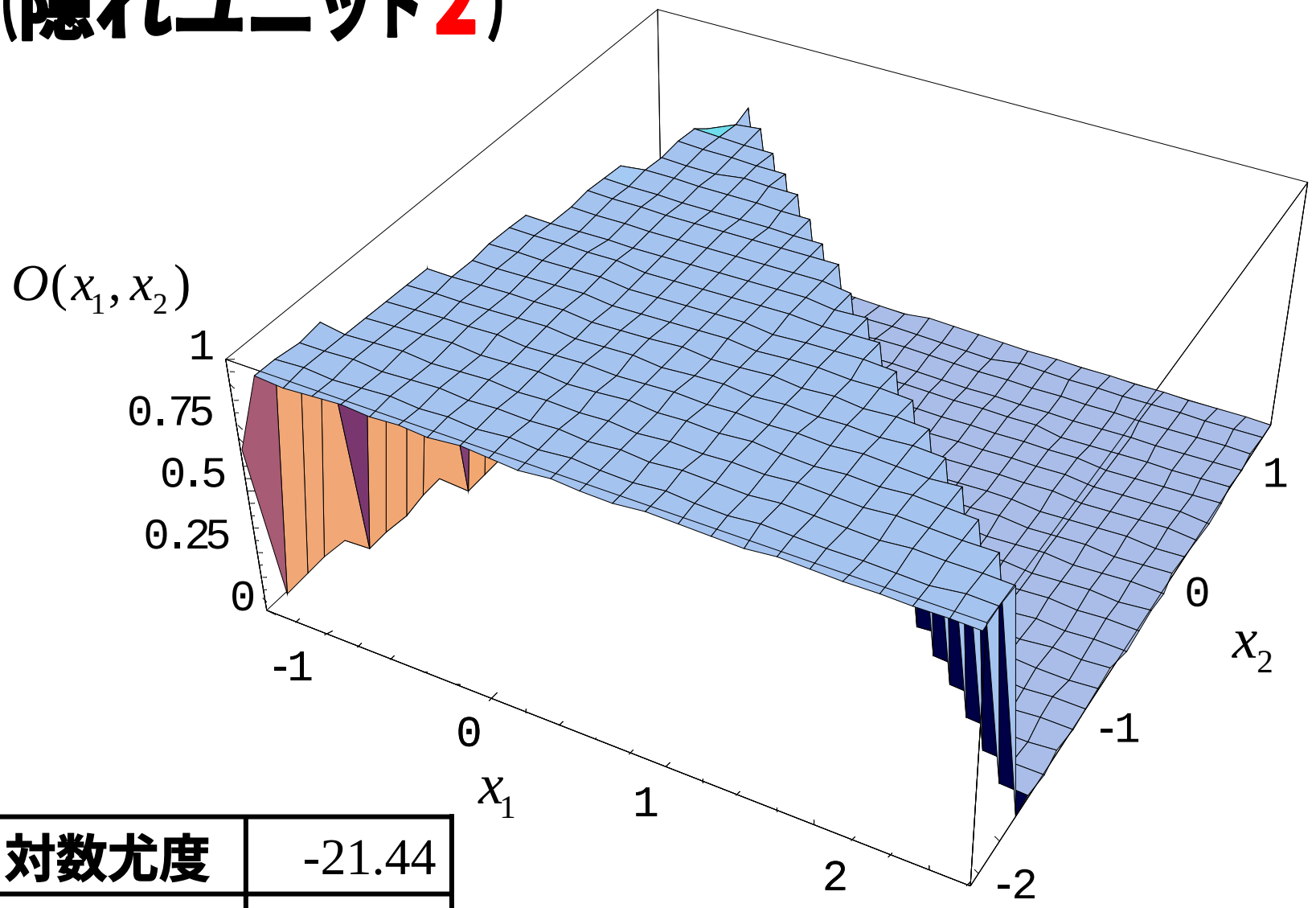
$$\begin{cases} o < 0.5 & : \text{(第1群　症状なし)} \\ o \geq 0.5 & : \text{(第2群　症状あり)} \end{cases}$$



対数尤度	-31.86
誤判別率	0.204

（x_1, x_2 は正規化された値）

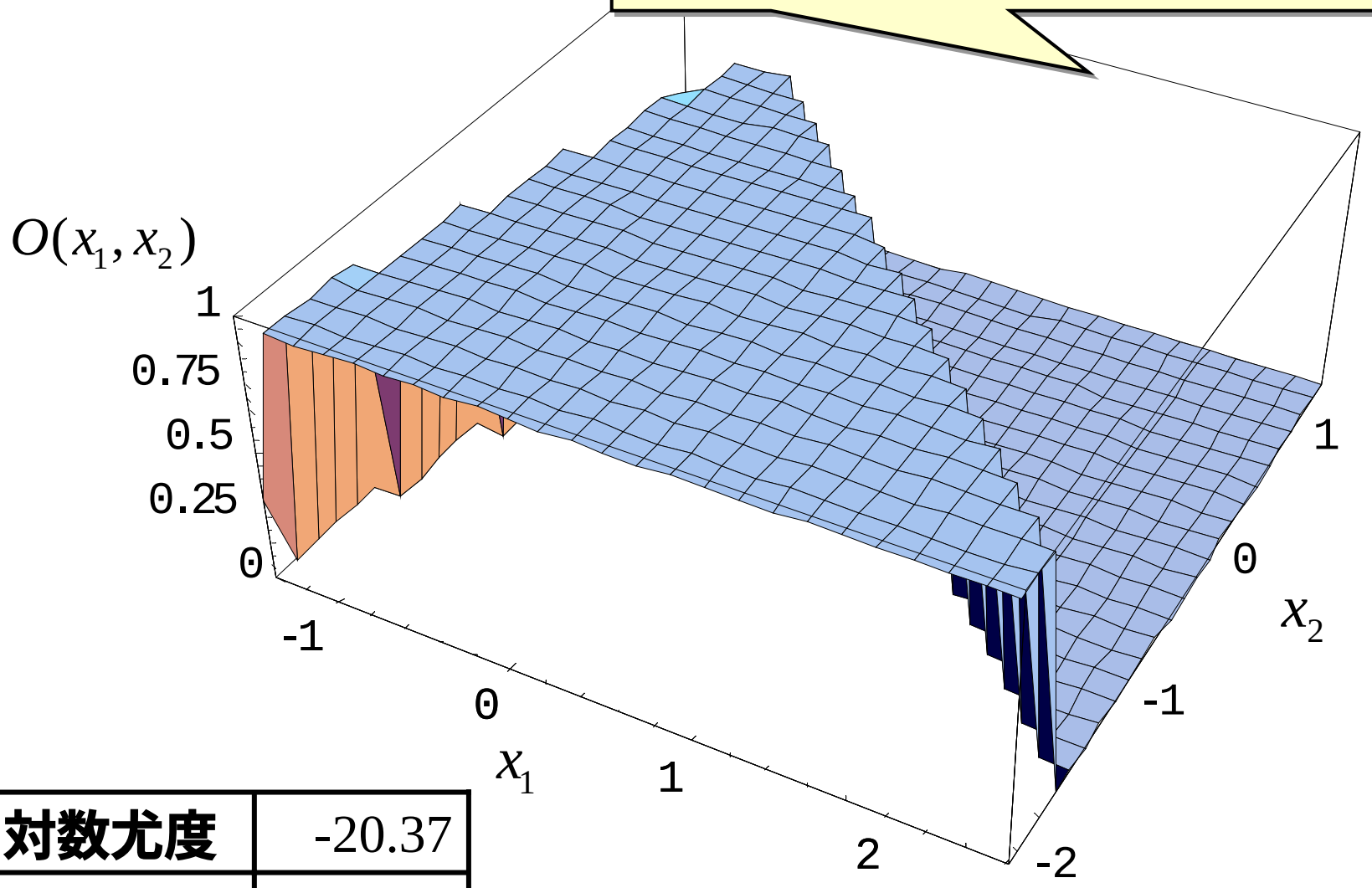
ニューロ判別モデルの出力
（隠れユニット2）



対数尤度	-21.44
誤判別率	0.120

ニューロ判別モデルの出力（隠れユニット3）

これが、本当に脊柱後湾症の症状あり、なしを分ける関数なのか？



対数尤度	-20.37
誤判別率	0.120

群を決定する関数を自分で設定

$$P_r = \frac{1}{1+\exp\{-f(x_1, x_2)\}} \qquad 0 \le P_r \le 1$$

$$f(x_1, x_2) = \sin(2\pi x_1) + x_1 x_2 + \sin(2\pi x_2)$$

$$\text{群（教師値 t）} = \begin{cases} t = 0 & \text{（第1群）} : P_r < 0.5 \\ t = 1 & \text{（第2群）} : P_r \ge 0.5 \end{cases}$$

$x_1 = 0.80, x_2 = 0.43$ **のとき,**

$$P_r(0.80, 0.43) = \frac{1}{1+\exp[-\sin(2\pi \times 0.80) - 0.80 \times 0.43 + \sin(2\pi \times 0.43)]}$$

$= 0.44$ → 0.5**より小さいので、第1群** (t=0)

関数の近似精度の評価

ニューロ判別モデル（2群判別）

- 出力と教師値の誤差より未知パラメータを推定
- 隠れユニット数を増やすことにより任意の関数を近似

$$o = \frac{1}{1 + \exp\left[-\sum_{j=0}^{J} \left\{ \frac{\beta_j}{1 + \exp\left(-\sum_{i=0}^{I} \alpha_{ij} x_i \right)} \right\} \right]}$$

比較不可能

比較可能

実際のデータ

- 群の決定は未知の関数

シミュレーションデータ

- 群を決定する関数を設定

2．シミュレーションの設計

X_1, X_2 ：[0,1]の一様乱数により生成．

シミュレーションデータ（訓練標本） No．1～1000

No.	X_1	X_2
1	0.80	0.43
2	0.77	0.85
3	0.64	0.45
4	0.73	0.40
5	0.55	0.40
6	0.66	0.68
：	：	：
：	：	：

No.	X_1	X_2
：	：	：
：	：	：
995	0.49	0.99
996	0.44	0.23
997	0.50	0.31
998	0.98	0.14
999	0.90	0.34
1000	0.50	0.22

シミュレーションデータの2群化

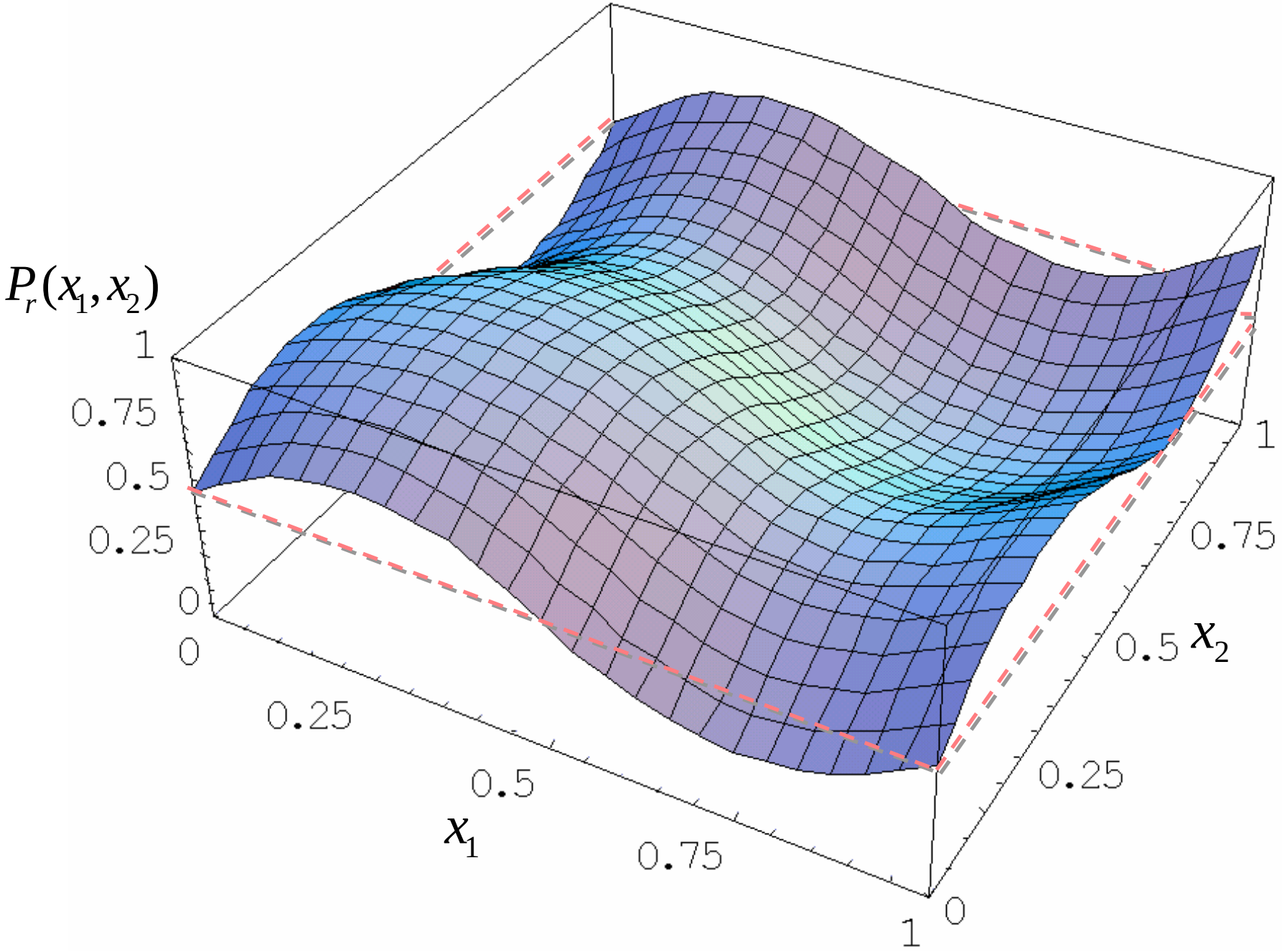
(群を決定する関数形の設定)

$$P_r = \frac{1}{1+\exp\{-f(x_1, x_2)\}} \qquad 0 \le P_r \le 1$$

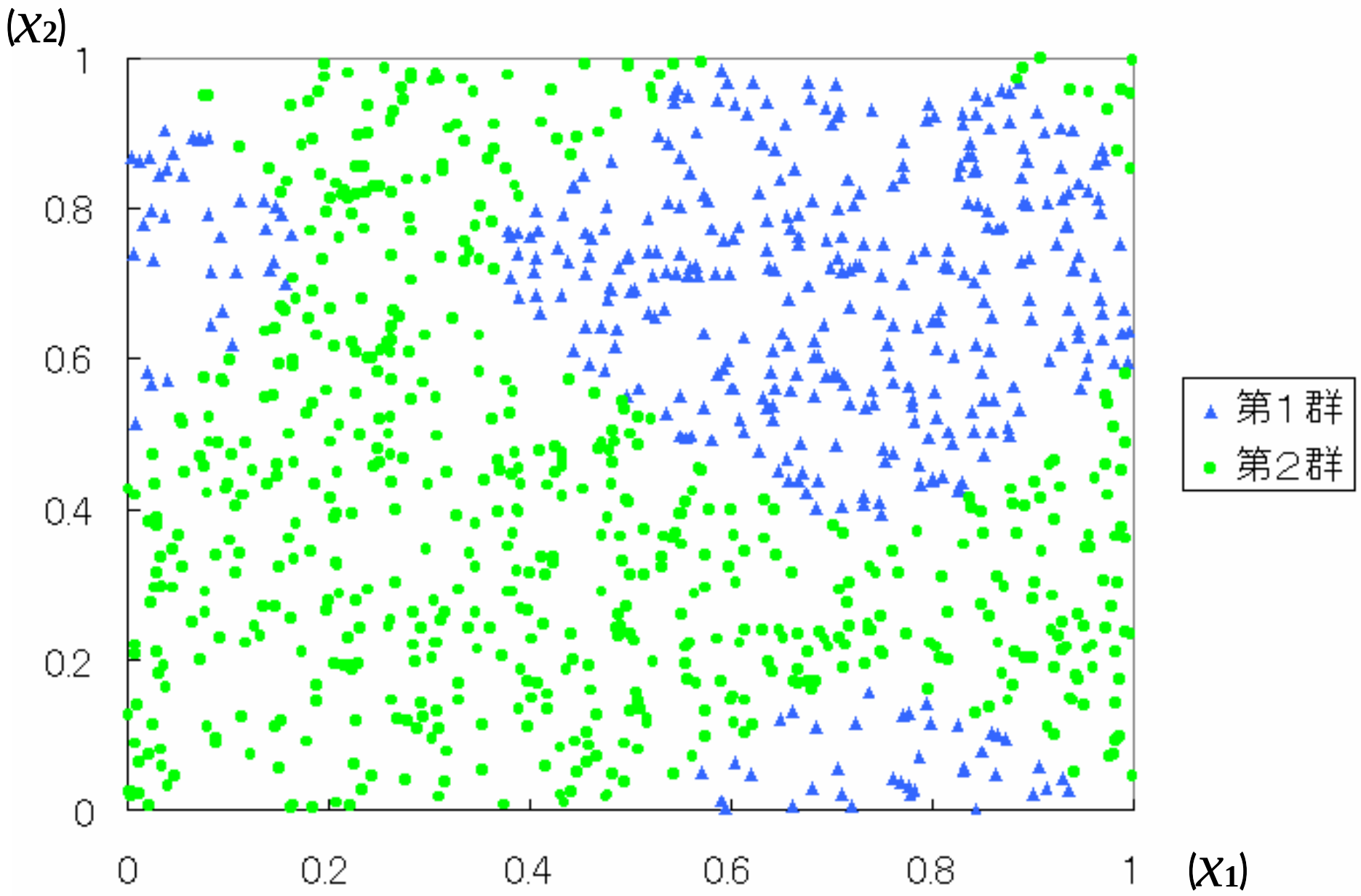
$$f(x_1, x_2) = \begin{cases} \sin(2\pi x_1) + x_1 x_2 + \sin(2\pi x_2) \\ \sin(2\pi x_1) + x_2 \\ \sin(2\pi x_1) + \sin(2\pi x_2) \end{cases}$$

群 (教師値 t) $= \begin{cases} t = 0 & \text{(第1群)} : P_r < 0.5 \\ t = 1 & \text{(第2群)} : P_r \ge 0.5 \end{cases}$

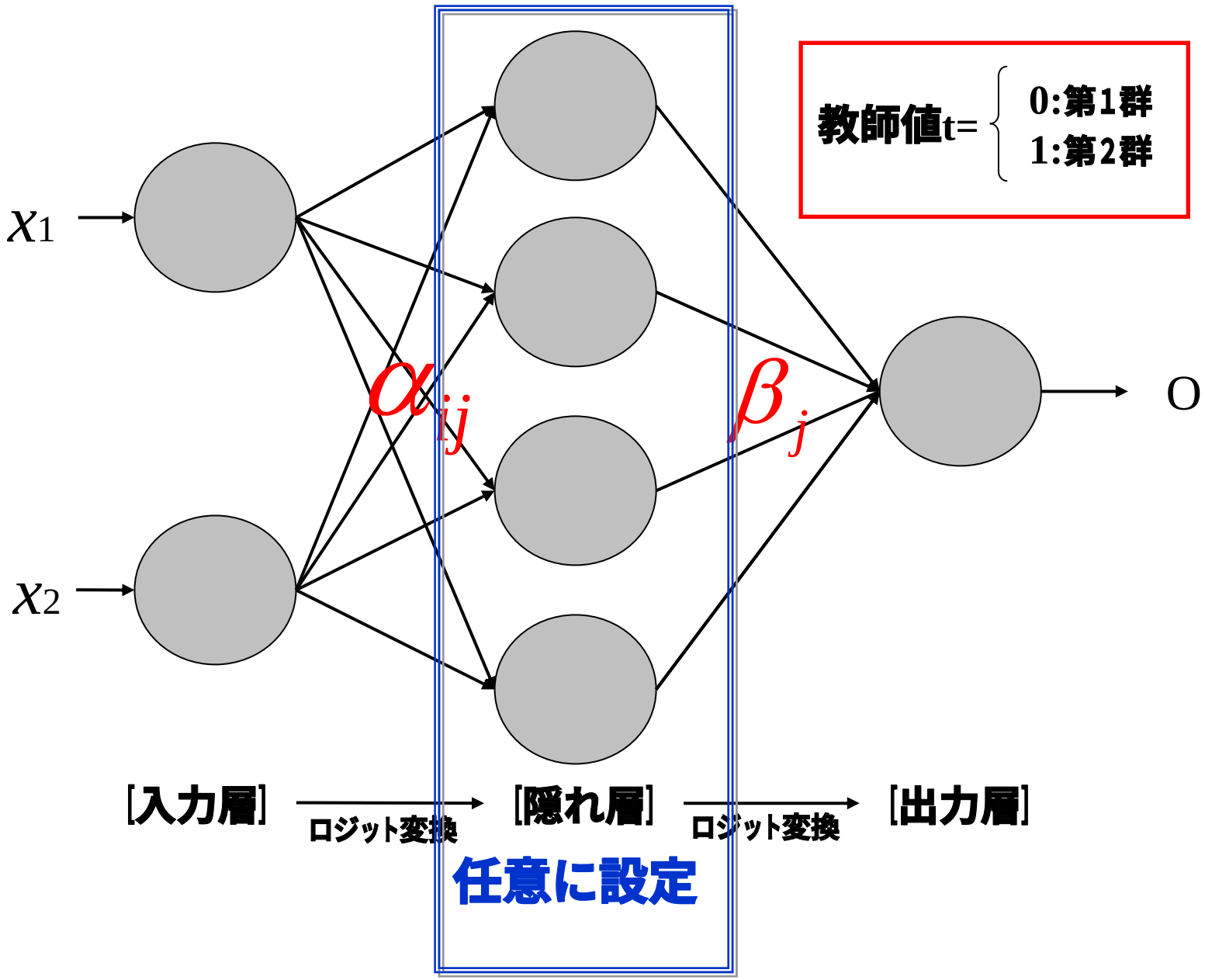
群を決定する関数形のイメージ



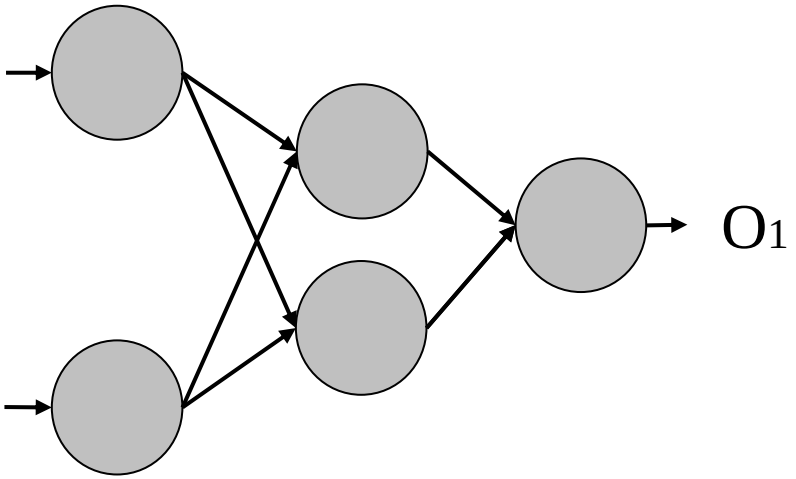
シミュレーションデータの群の分布



3．ニューロ判別モデルの構築

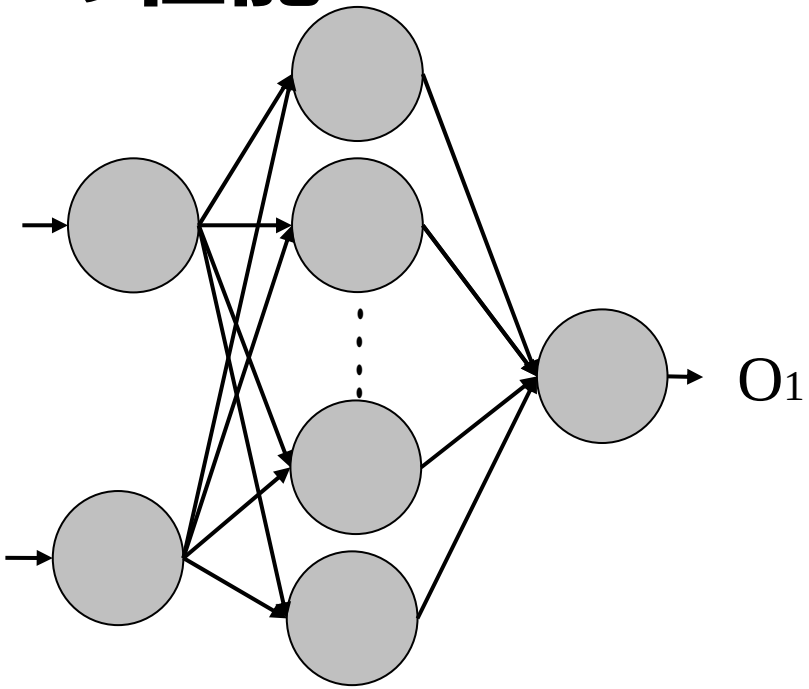


隠れユニット数と，モデルの性能



[隠れユニット数が少なすぎる場合]

・十分な学習ができず,データ構造を適切に表現できないため,本質的な関係を捉えることができない.



[隠れユニット数が多すぎる場合]

・過学習 (Over fitting) ,モデルの複雑化が起り,偶然変動に過敏に反応し,かえって本質的な関係を見失う.

最適な隠れユニット数の決定による，
妥当なネットワークの構築が望まれる.

隠れユニット数決定に用いる情報量規準

AIC (*A*kaike *I*nformation *C*riterion)

$$AIC = -2\ln L\left(\boldsymbol{X};\hat{\theta}(\boldsymbol{X})\right)+2p$$

p：モデルの自由なパラメータ数
（モデルのバイアスを近似）

EIC (*E*xtended *I*nformation *C*riterion)

$$EIC = -2\ln L\left(\boldsymbol{X};\hat{\theta}(\boldsymbol{X})\right)+2C^{*}$$

C^{*}＝ブートストラップ･バイアス推定量
（モデルのバイアスを数値的に算出）

ブートストラップ法による，バイアスの推定法

手順1　訓練標本から，リサンプリングによりブートストラップ標本を生成.

$$X = \{X_1, X_2, X_3, \cdots\cdots, X_{998}, X_{999}, X_{1000}\}$$

[訓練標本]

$$X^* = \{X^*_1, X^*_2, X^*_3, \cdots\cdots, X^*_{998}, X^*_{999}, X^*_{1000}\}$$

[ブートストラップ標本]

生成したブートストラップ標本より,ニューロ判別モデルの構築.

手順2　対数尤度の算出.

$$\ln L\left(X^*;\hat{\theta}\left(X^*\right)\right) \qquad \theta=\left\{\ \alpha_{ij},\beta_j\ \right\}$$

～ブートストラップ標本より構築したニューラルネットワークの対数尤度.

$$\ln L\left(X;\hat{\theta}\left(X^*\right)\right)$$

～ブートストラップ標本より構築したニューラルネットワークに元の訓練標本を当てはめたときの対数尤度.

手順3　手順1，2を必要回数繰り返す．(本実験ではB=200回)

手順4　手順3で得られた値をより、バイアスのブートストラップ推定.

$$C^* \approx \frac{1}{B}\sum_{b=1}^{B}\left\{\ln L\left(X_b^*;\hat{\theta}\left(X_b^*\right)\right)-\ln L\left(X;\hat{\theta}\left(X_b^*\right)\right)\right\}$$

ブートストラップ法に基づく，情報量規準.

$$EIC=-2\ln L\left(X;\hat{\theta}\left(X\right)\right)+2C^* \quad \Rightarrow \quad \textbf{最小化}$$

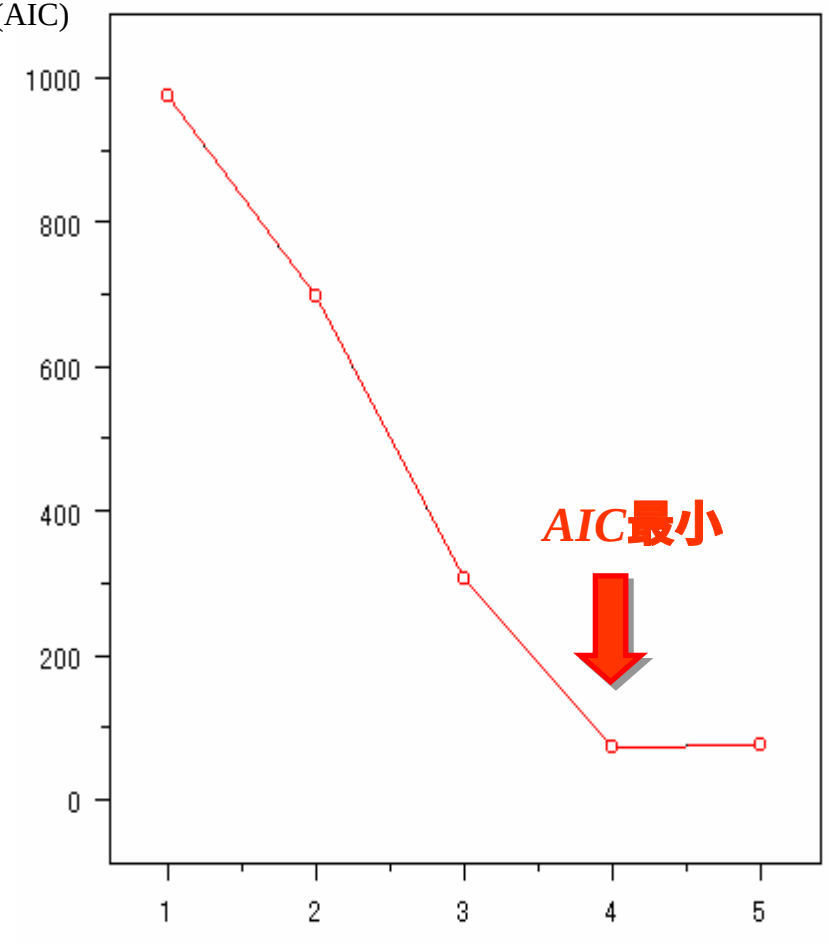
もとの対数尤度

バイアスのブートストラップ推定

*AIC*と*EIC*の挙動

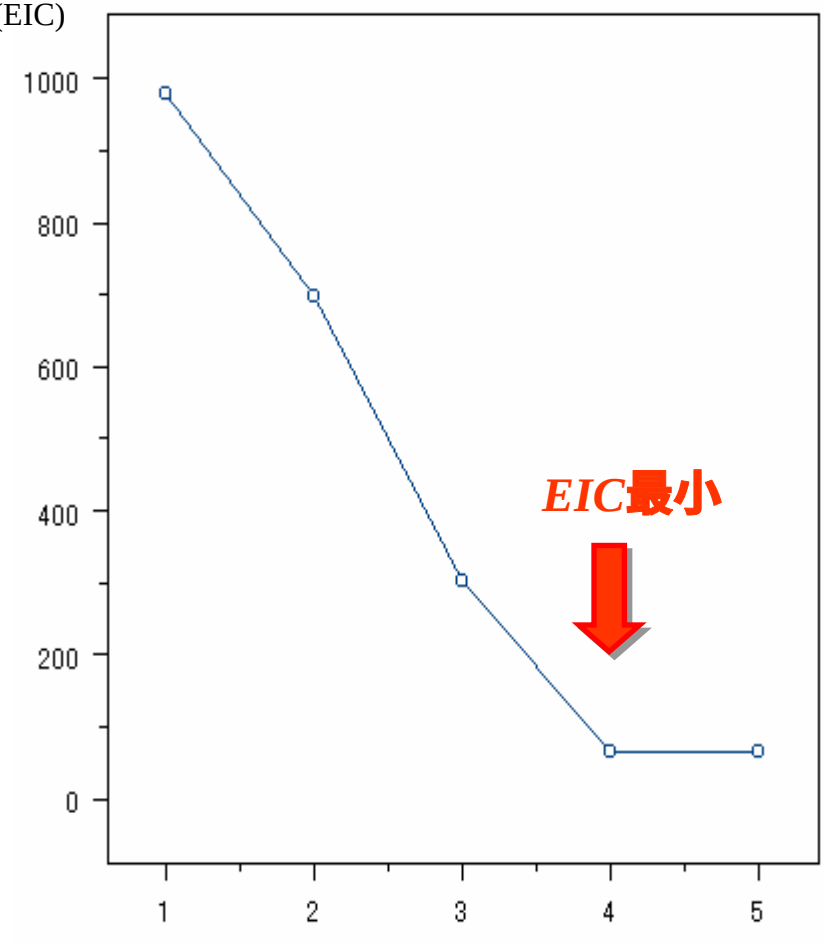
隠れユニット数	1	2	3	4	5
対数尤度	-481.82	-339.62	-140.59	-19.72	-16.24
AIC	973.65	697.25	307.19	73.43	74.47
EIC	978.72	697.90	303.94	64.25	66.10

(AIC)



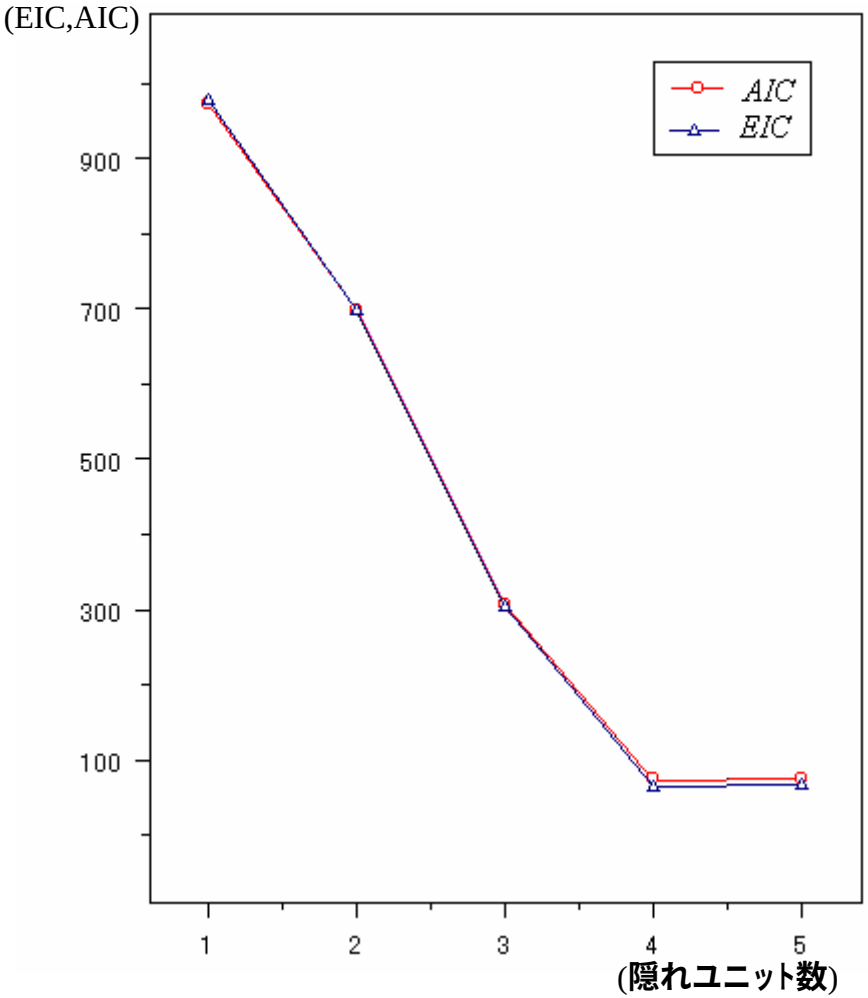
(隠れユニット数)

(EIC)



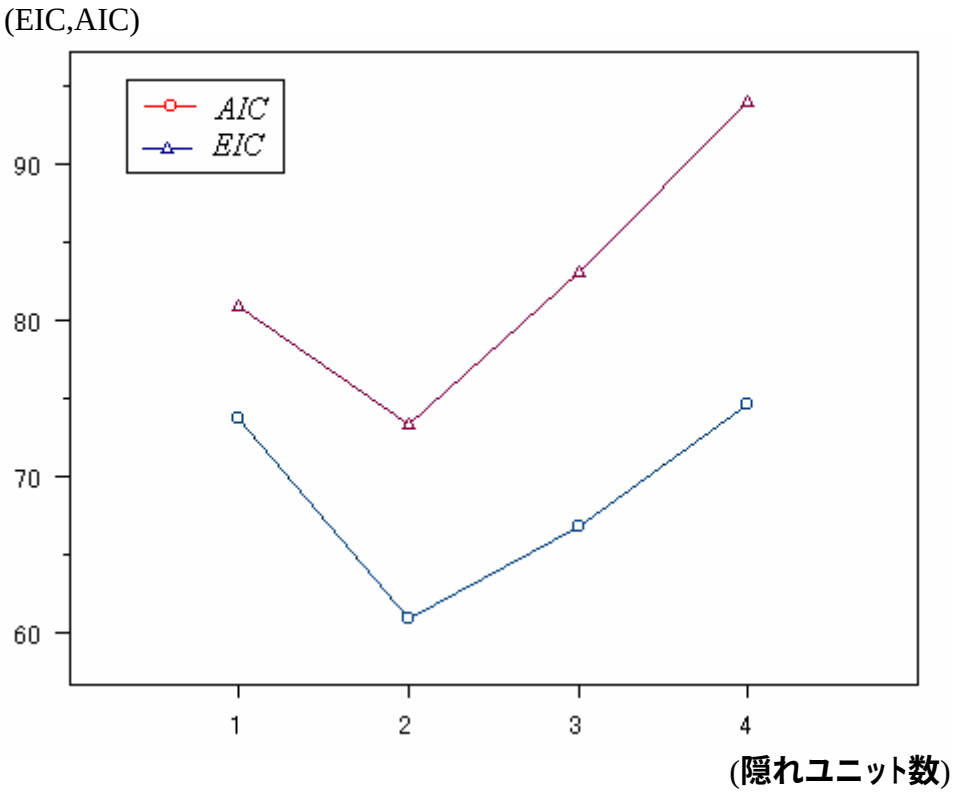
(隠れユニット数)

*AIC*と*EIC*の挙動 (2)



シミュレーションデータ

AICは想定したモデル族に真のモデルが含まれることを前提にしている.



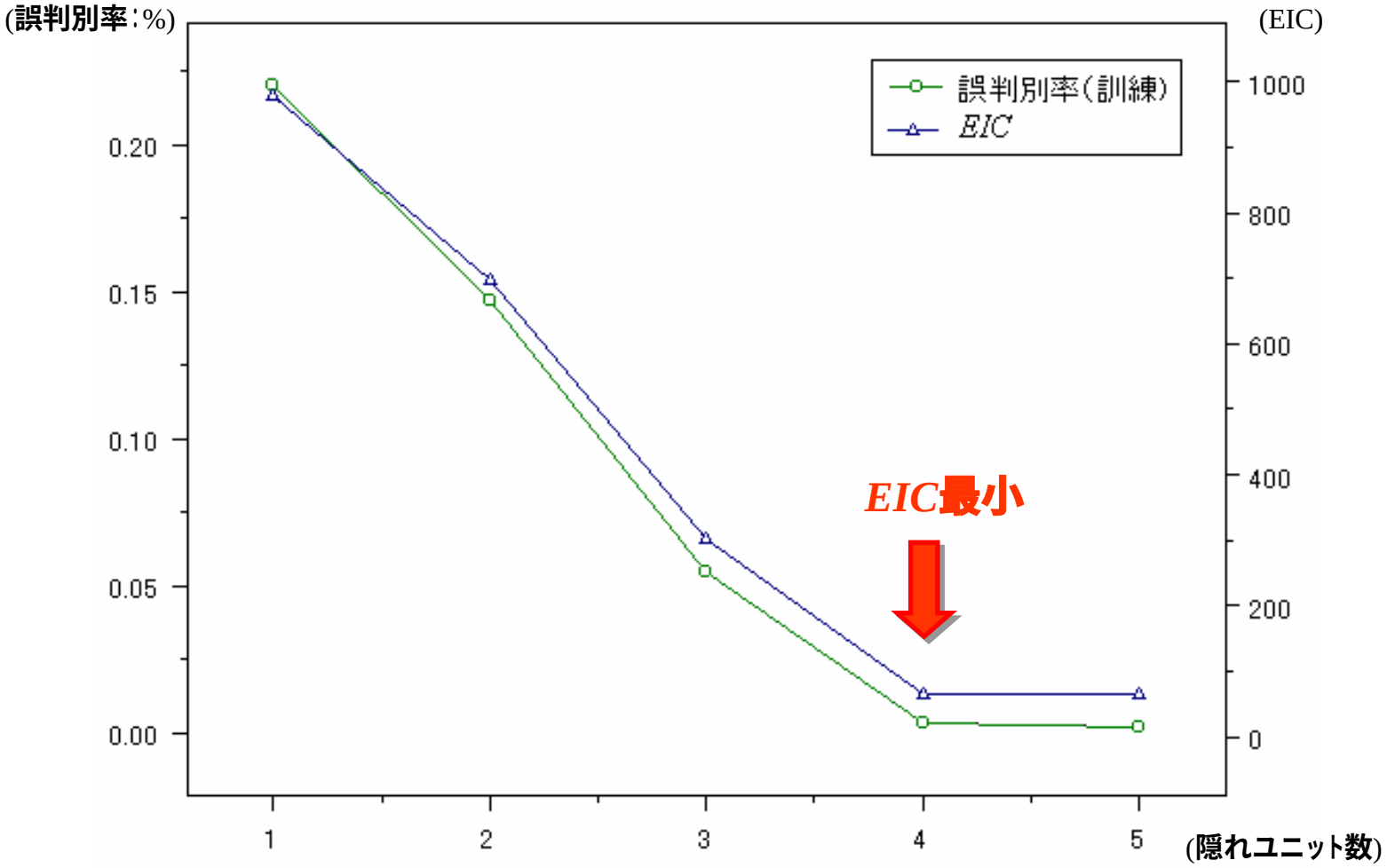
AICとEICが,ほぼ一致する.

実際のデータの場合では ⇒

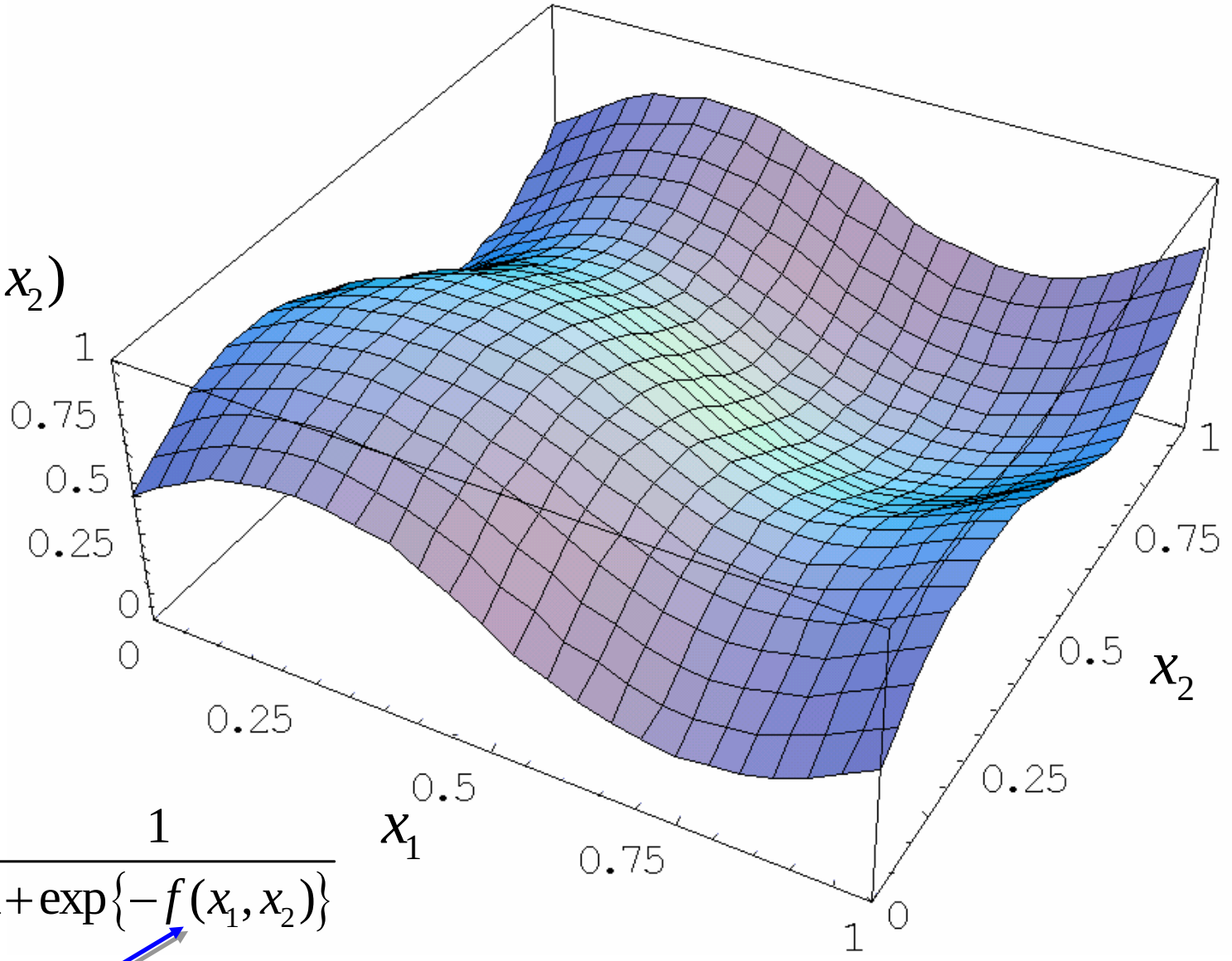
実際のデータ (初めに紹介した脊柱後湾症の解析)

誤判別率 (訓練標本)

隠れユニット数	1	2	3	4	5
対数尤度	-481.82	-339.62	-140.59	-19.72	-16.24
EIC	978.72	697.90	303.94	64.25	66.10
誤判別率 (訓練)	0.220	0.147	0.055	0.003	0.002



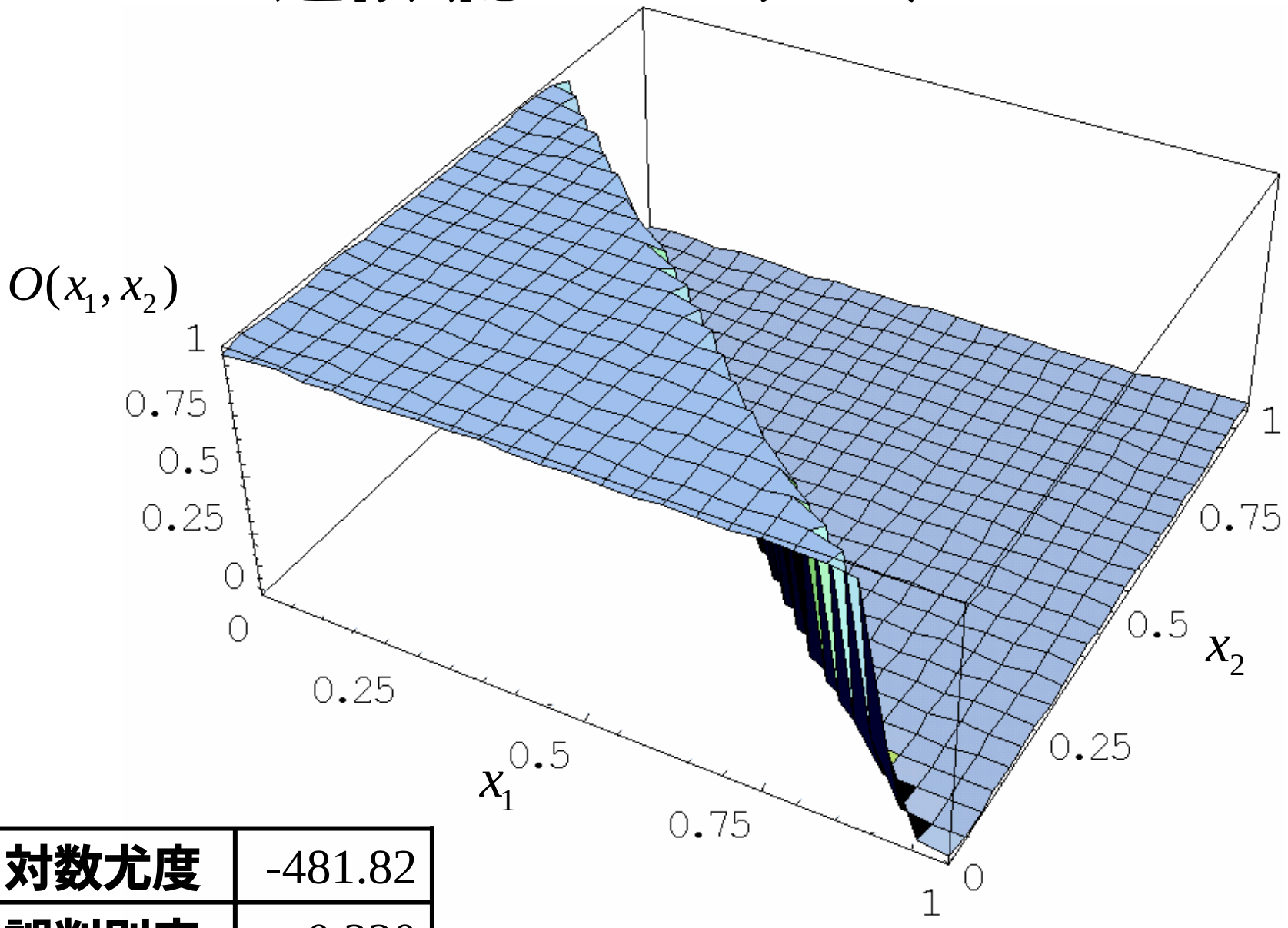
群を決定する関数形のイメージ



$$P_r = \frac{1}{1+\exp\{-f(x_1, x_2)\}}$$

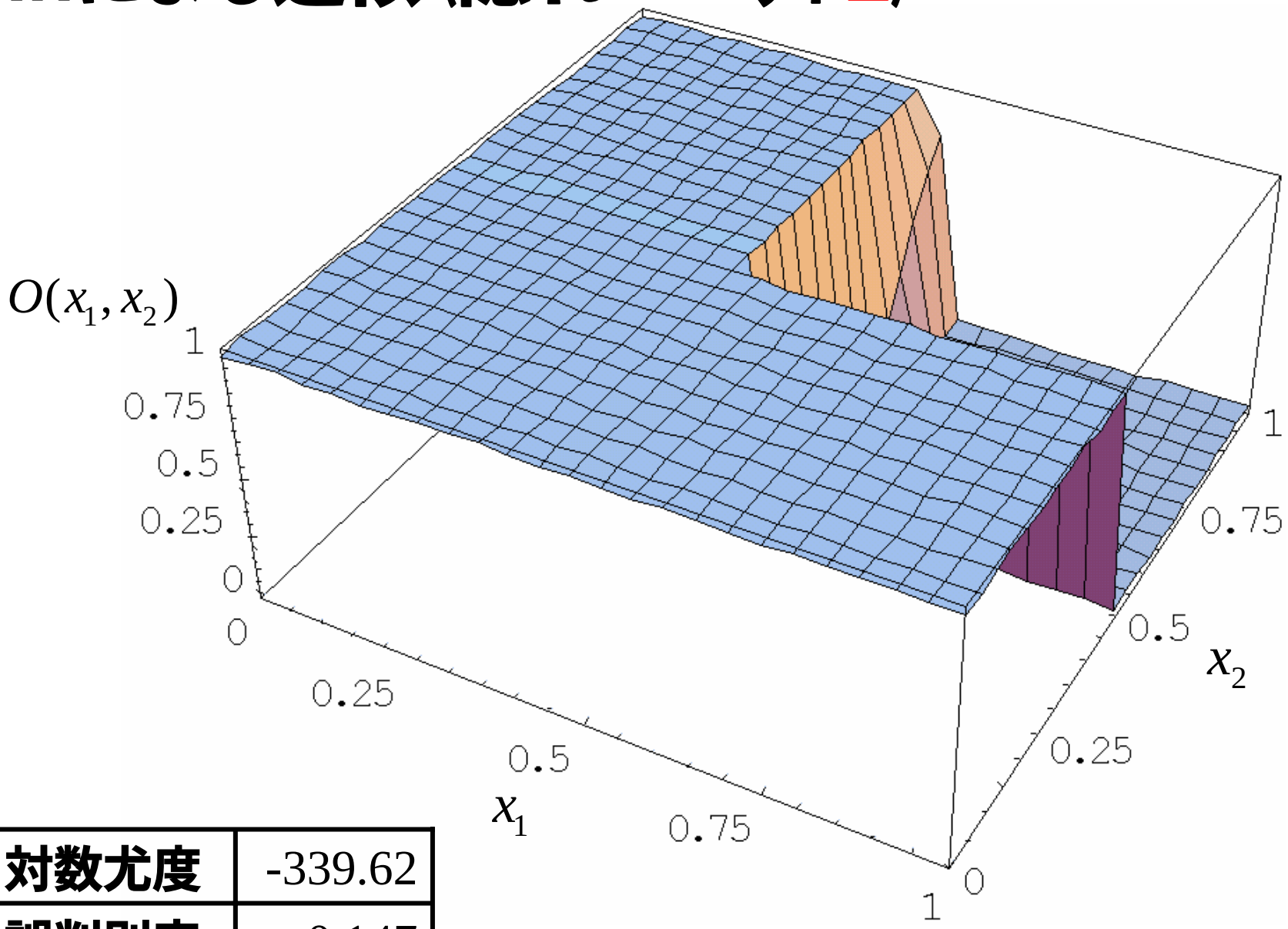
$$f(x_1, x_2) = \sin(2\pi x_1) + x_1 x_2 + \sin(2\pi x_2)$$

NNによる近似 (隠れユニット1)



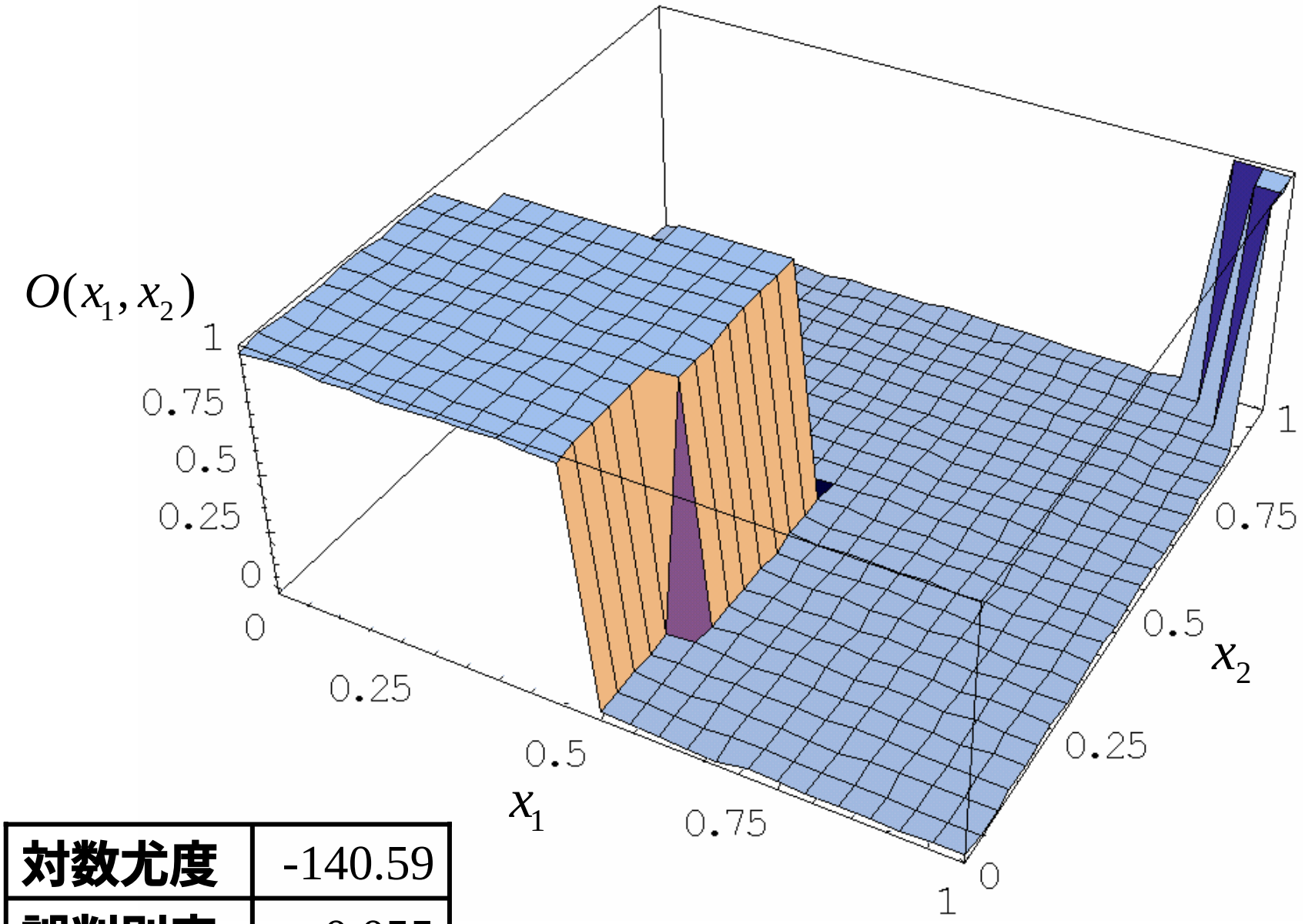
対数尤度	-481.82
誤判別率	0.220

NNによる近似 (隠れユニット2)



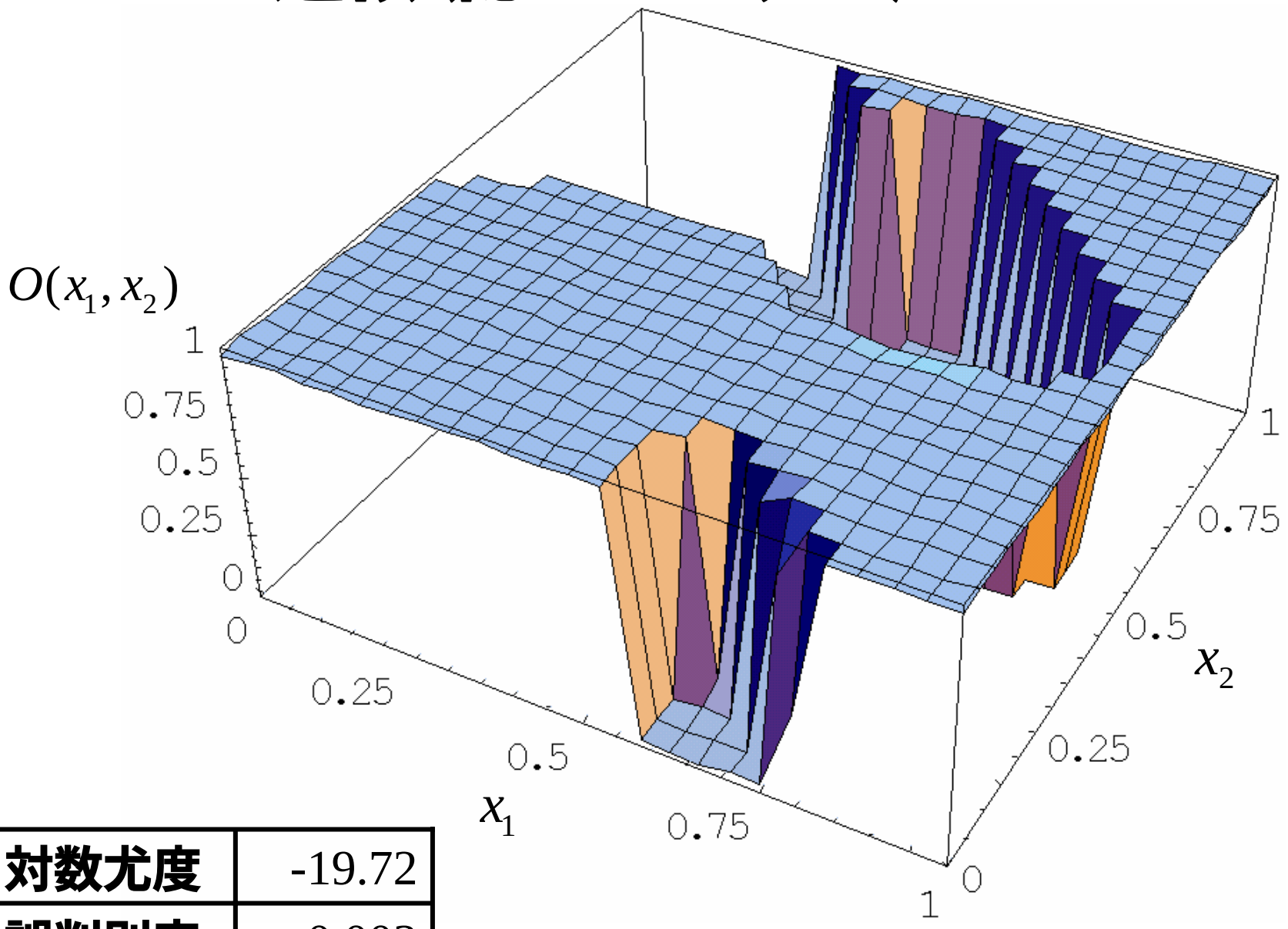
対数尤度	-339.62
誤判別率	0.147

NNによる近似 (隠れユニット3)



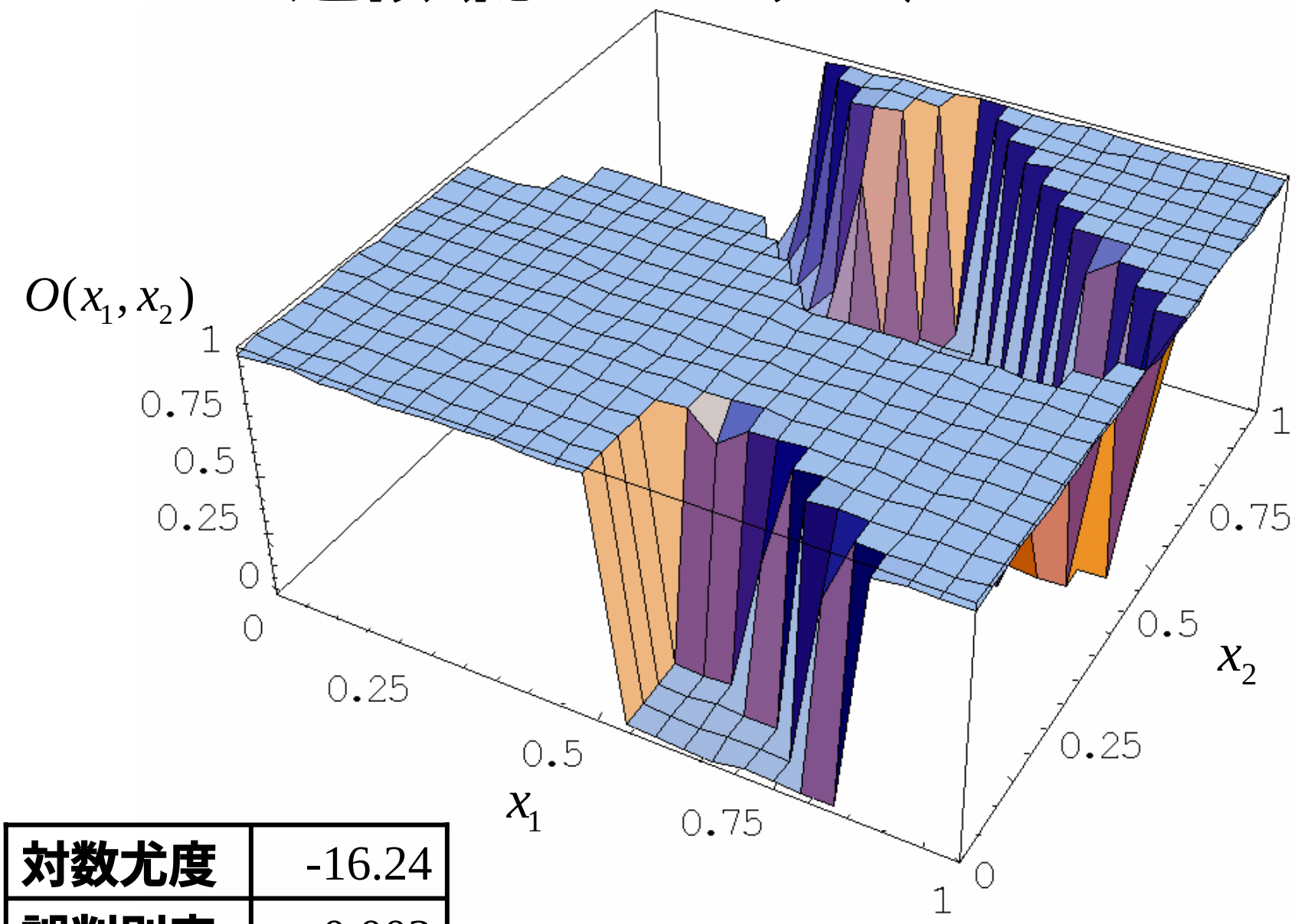
対数尤度	-140.59
誤判別率	0.055

NNによる近似 (隠れユニット4)



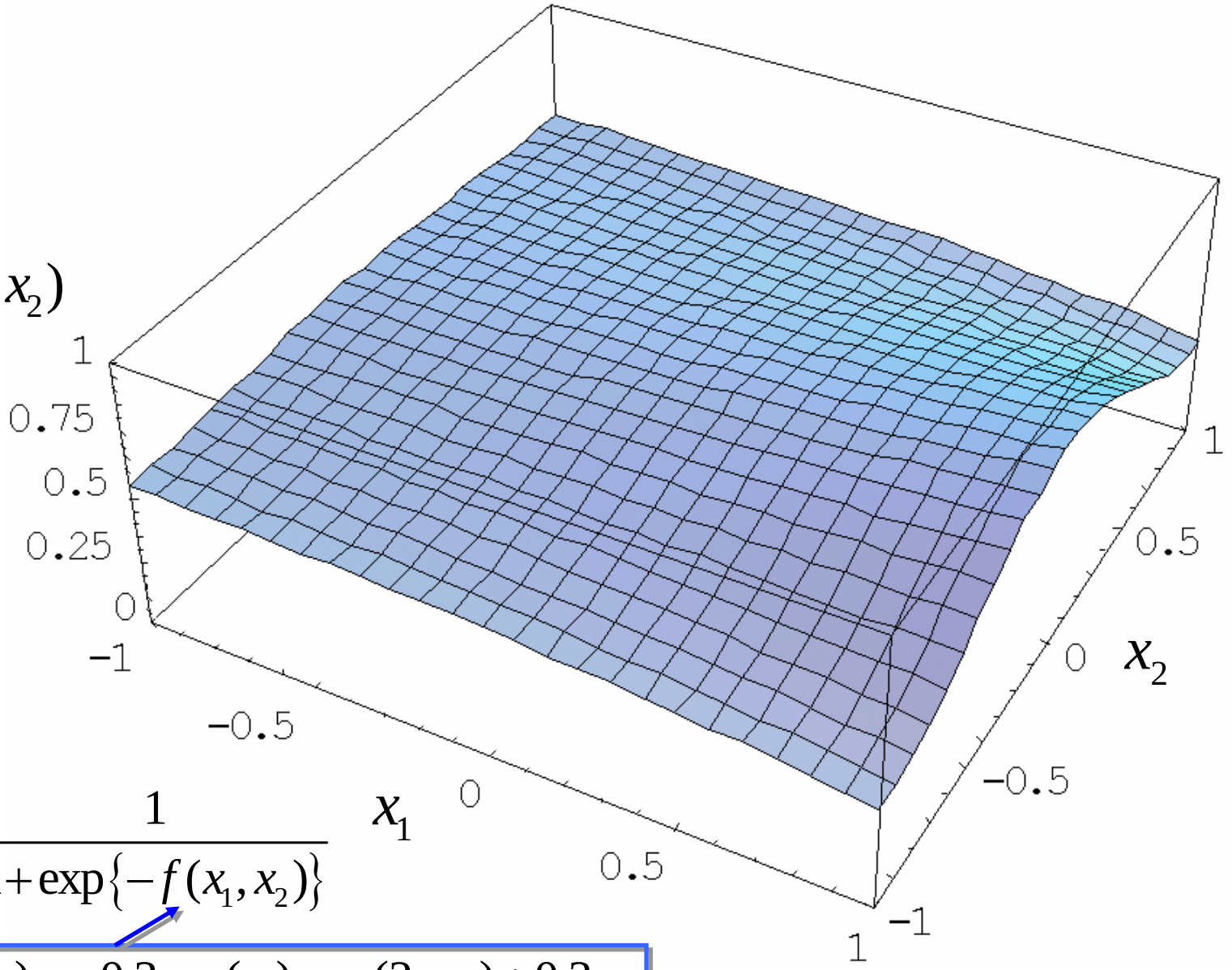
対数尤度	-19.72
誤判別率	0.003

NNによる近似 (隠れユニット5)



対数尤度	-16.24
誤判別率	0.002

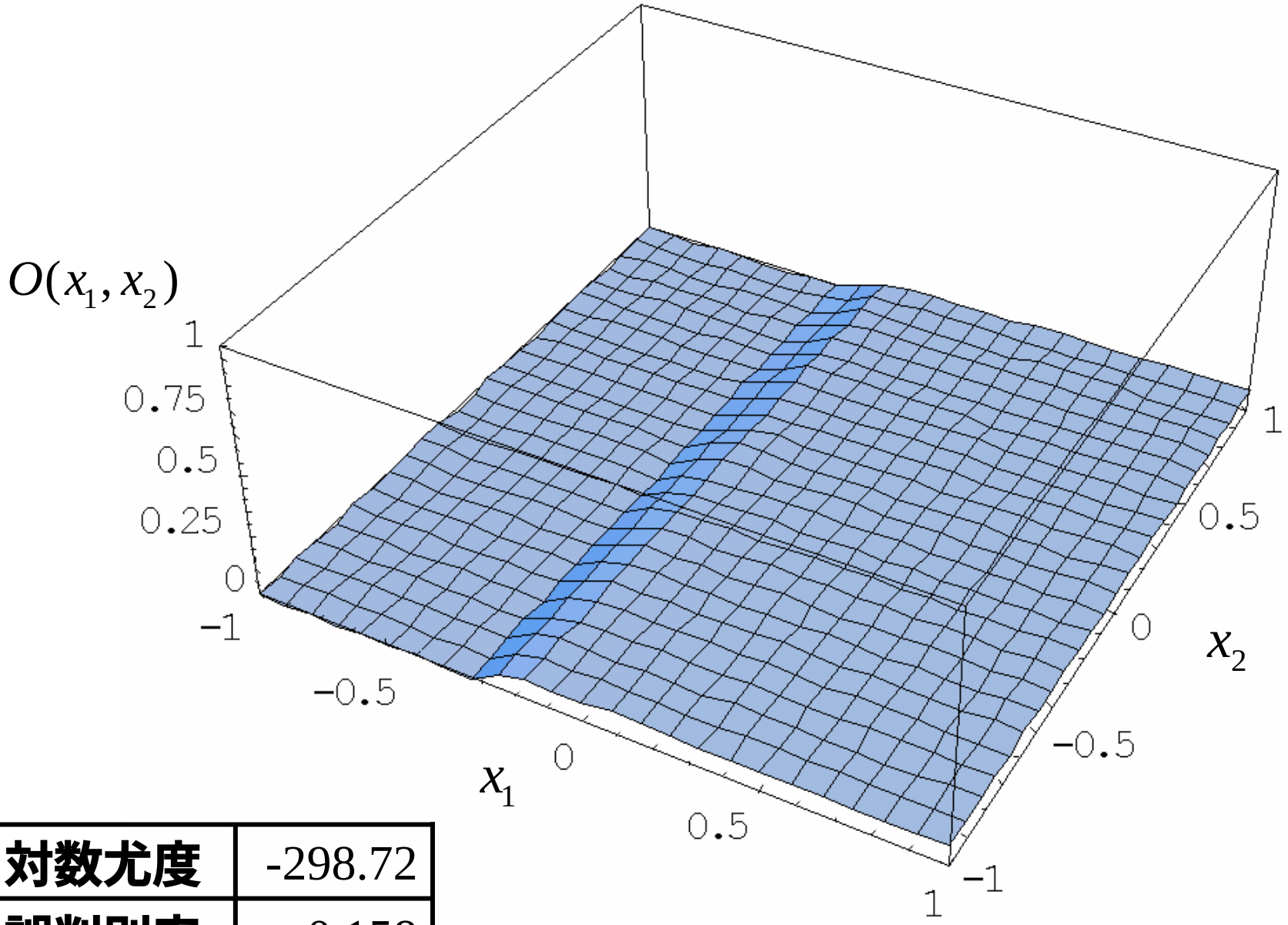
別の関数の場合　（X_1, X_2 ：[-1,1]の一様乱数）



$$P_r = \frac{1}{1+\exp\{-f(x_1, x_2)\}}$$

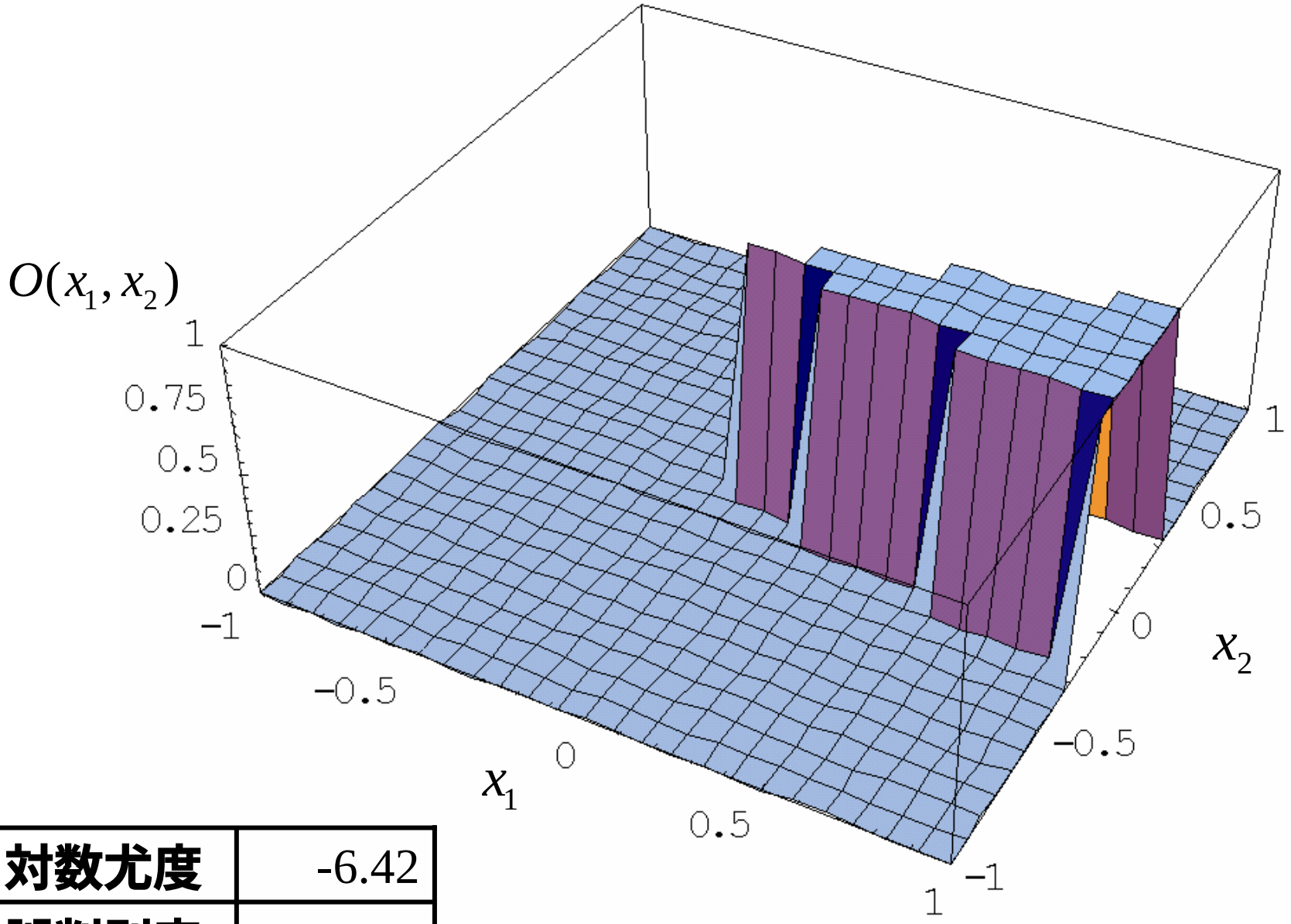
$$f(x_1, x_2) = -0.3\exp(x_1)\cdot\cos(2\pi x_2) + 0.3$$

NNによる近似（隠れユニット1）　AIC=607.45



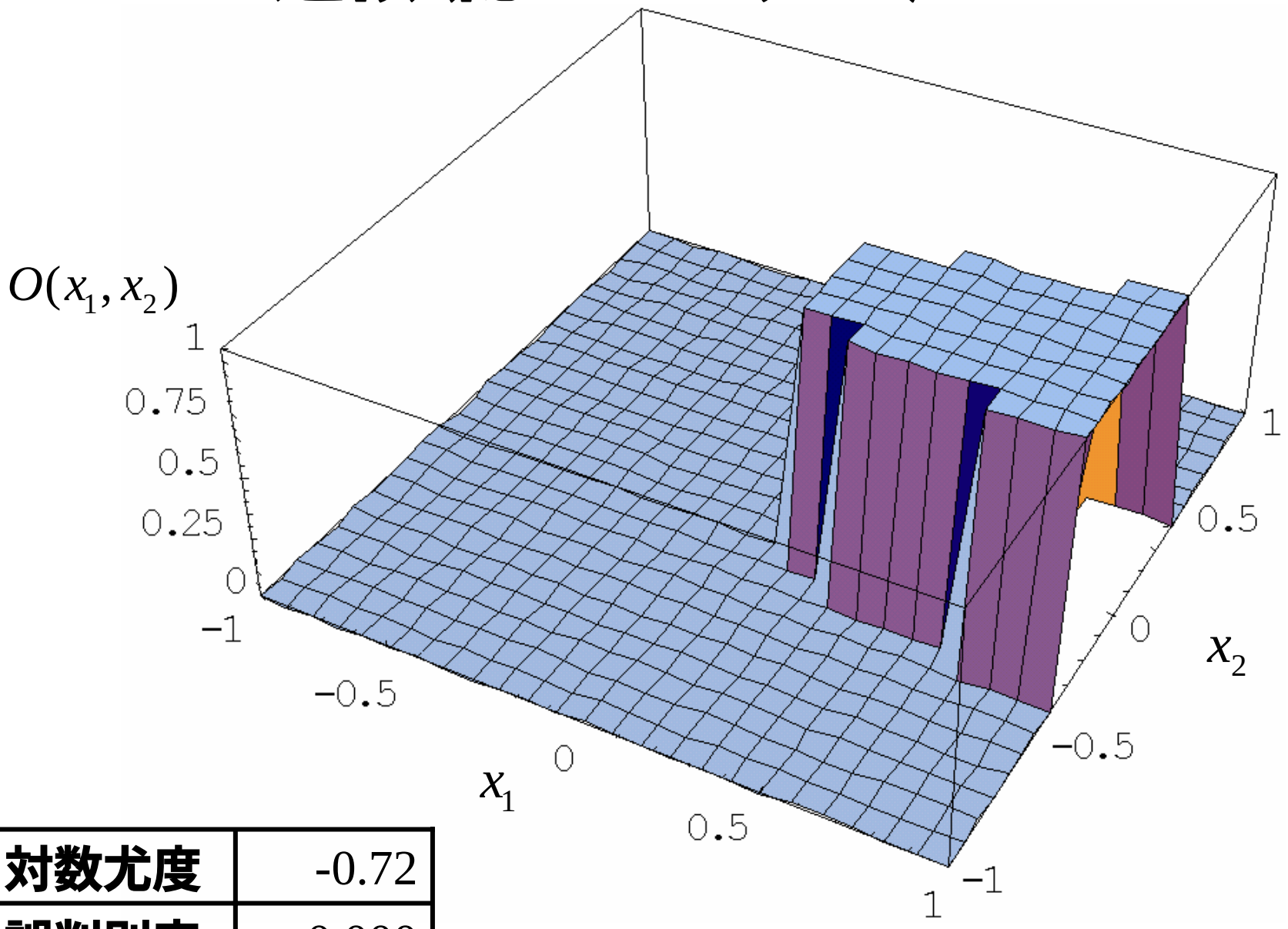
対数尤度	-298.72
誤判別率	0.158

NNによる近似（隠れユニット2）　AIC=30.85



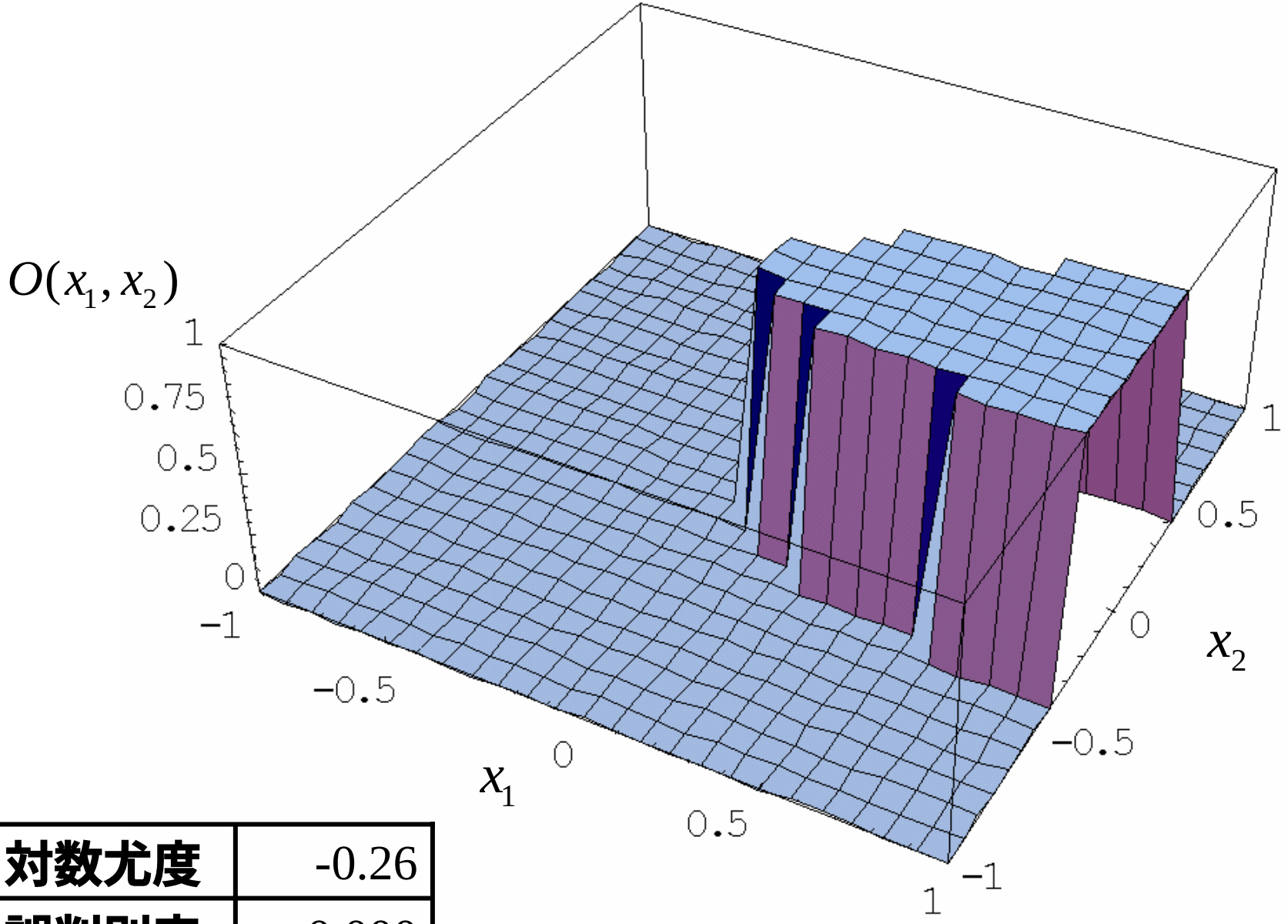
対数尤度	-6.42
誤判別率	0.003

NNによる近似（隠れユニット3）　AIC=27.44



対数尤度	-0.72
誤判別率	0.000

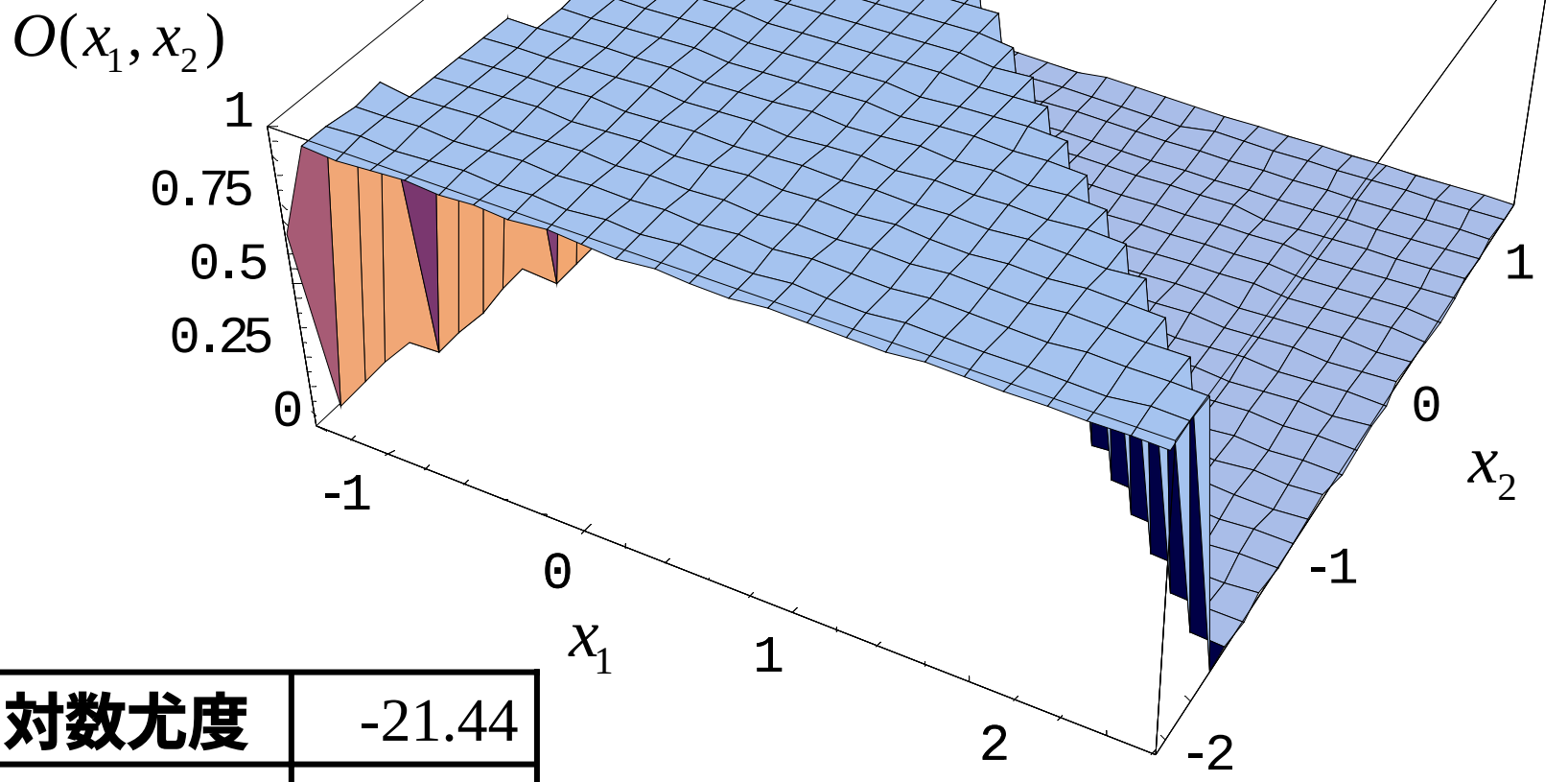
NNによる近似 (隠れユニット4)　AIC=34.53



対数尤度	-0.26
誤判別率	0.000

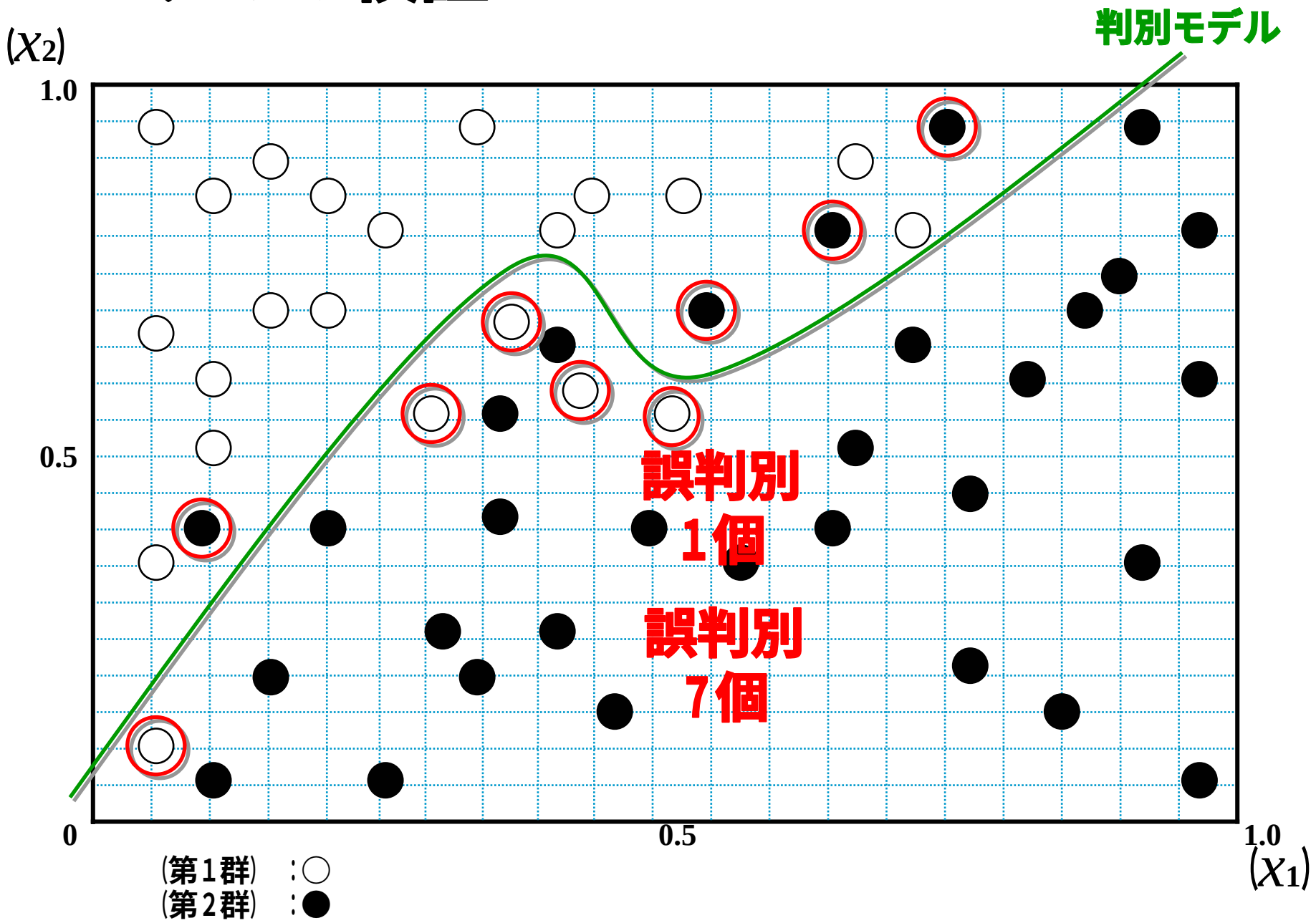
ニューロ判別モデルの出力（隠れユニット2）

これは脊柱後湾症の症状あり、なしを分ける関数であると考えることができそうである.



対数尤度	-21.44
誤判別率	0.120

4．モデルの検証



検証標本の生成

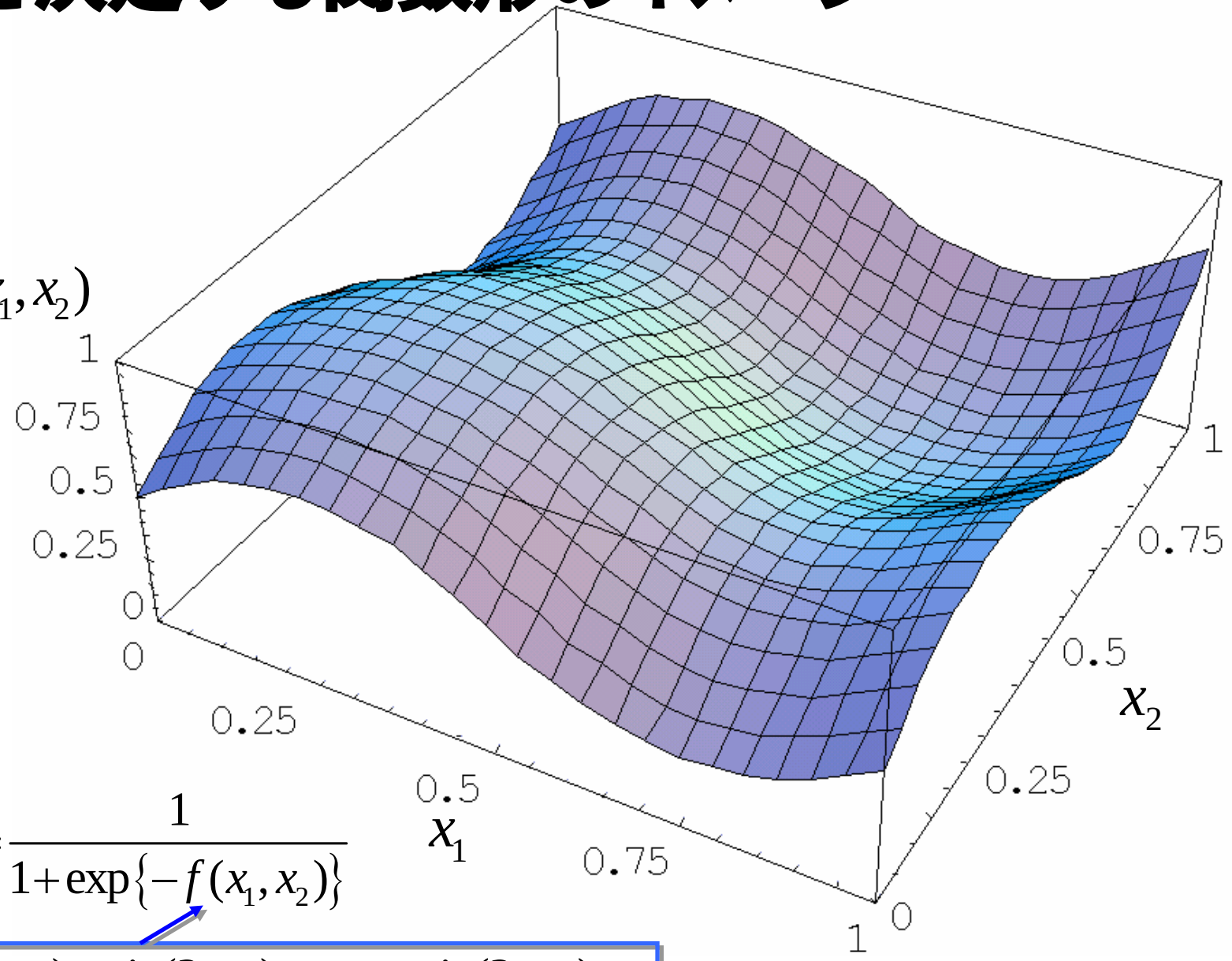
訓練標本で構築したモデルの検証の為に,訓練標本とは別の一様乱数を用いて、検証標本を生成.

シミュレーションデータ (検証標本)　No. 1～1000

No.	x_1	x_2
1	0.91	0.13
2	0.70	0.93
3	0.97	0.36
4	0.45	0.42
5	0.05	0.32
6	0.32	0.80
⋮	⋮	⋮
⋮	⋮	⋮

No.	x_1	x_2
⋮	⋮	⋮
⋮	⋮	⋮
995	0.45	0.09
996	0.82	0.51
997	0.94	0.61
998	0.30	0.41
999	0.64	0.57
1000	0.70	0.86

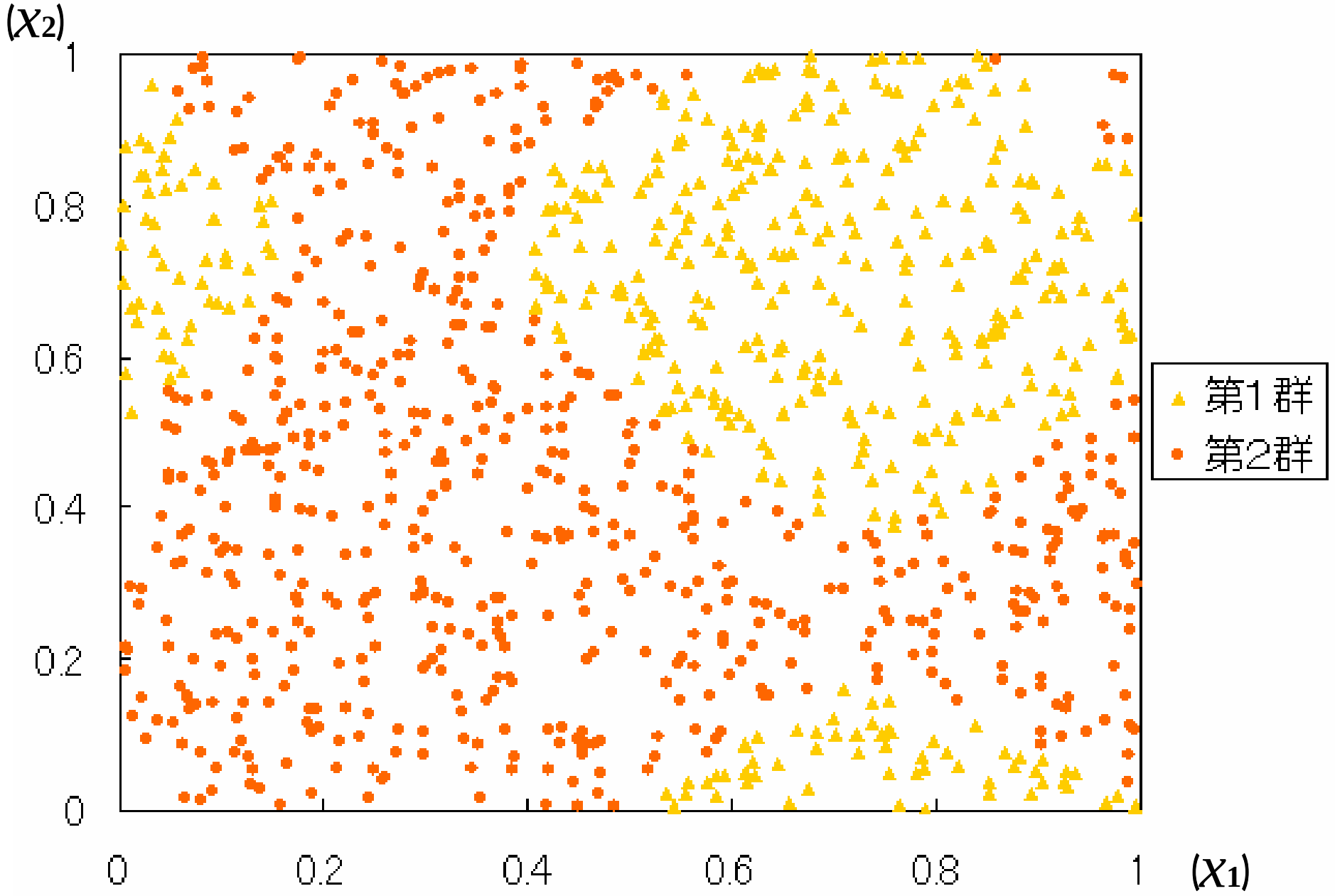
群を決定する関数形のイメージ



$$P_r = \frac{1}{1+\exp\{-f(x_1, x_2)\}}$$

$$f(x_1, x_2) = \sin(2\pi x_1) + x_1 x_2 + \sin(2\pi x_2)$$

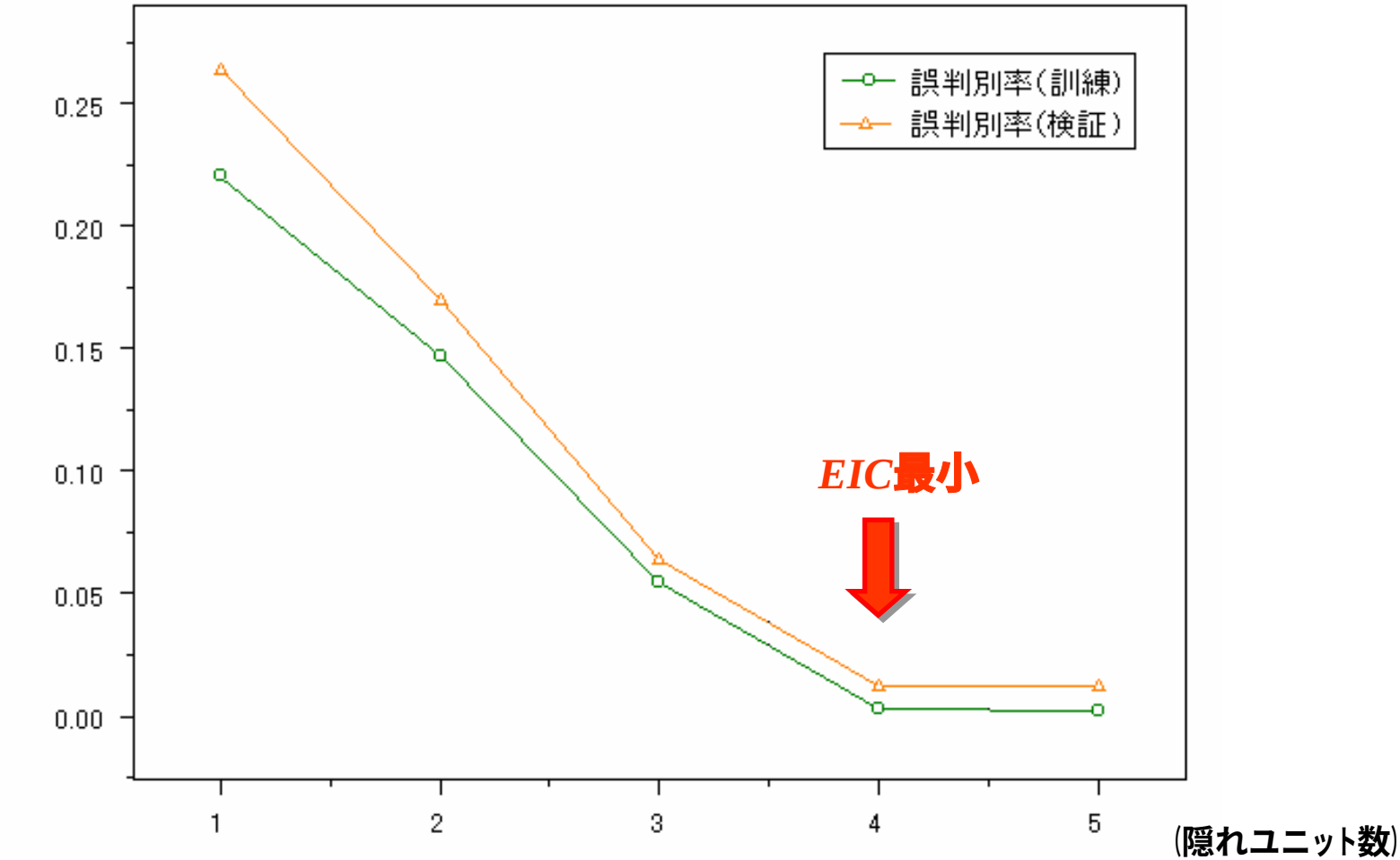
検証標本の群の分布



誤判別率の比較によるモデルの検証

隠れユニット数	1	2	3	4	5
誤判別率 (訓練)	0.220	0.147	0.055	0.003	0.002
誤判別率 (検証)	0.264	0.170	0.064	0.012	0.010

(誤判別率)



(隠れユニット数)

5．他のモデルとの性能比較

	誤判別率 (訓練)	誤判別率 (検証)
ニューロ判別モデル (隠れユニット数4)	0.003	0.012
線形判別 (2次判別)	?	?
ロジスティック判別	?	?

⇒訓練標本,検証標本に対する誤判別率を比較

フィッシャーの線形判別

線形判別関数

$$z = 4.33 - 3.82x_1 - 4.32x_2$$

($z < 0$ なら第1群, $z > 0$ なら第2群)

	誤判別率 (訓練)	誤判別率 (検証)
線形判別	0.234	0.266

2群間の等分散性は成立たない.

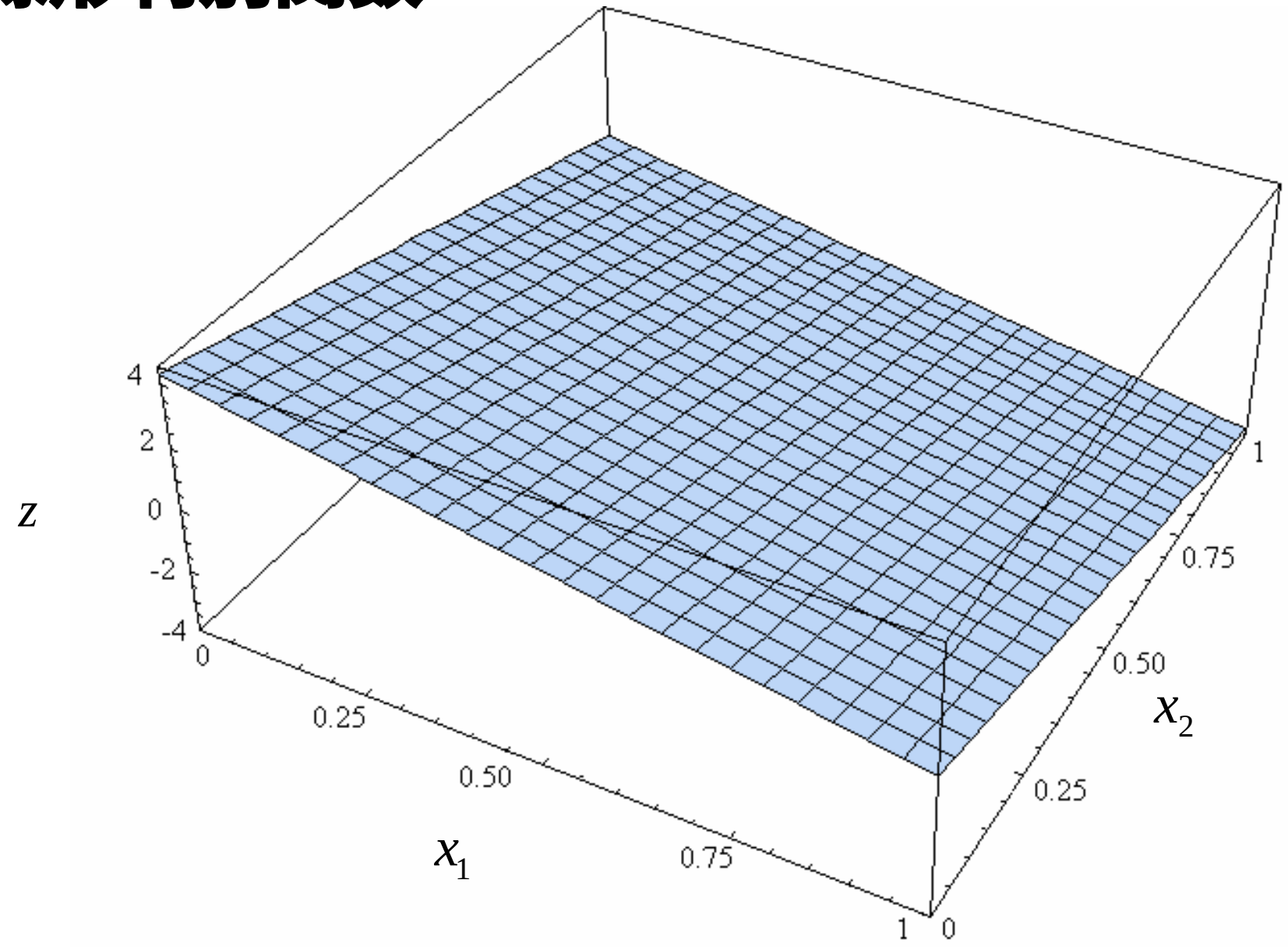
2次判別

2次判別関数

$$z = 12.85 - 15.38x_1 - 14.95x_2 + 5.04x_1^2 + 3.26x_1x_2 + 3.90x_2^2$$

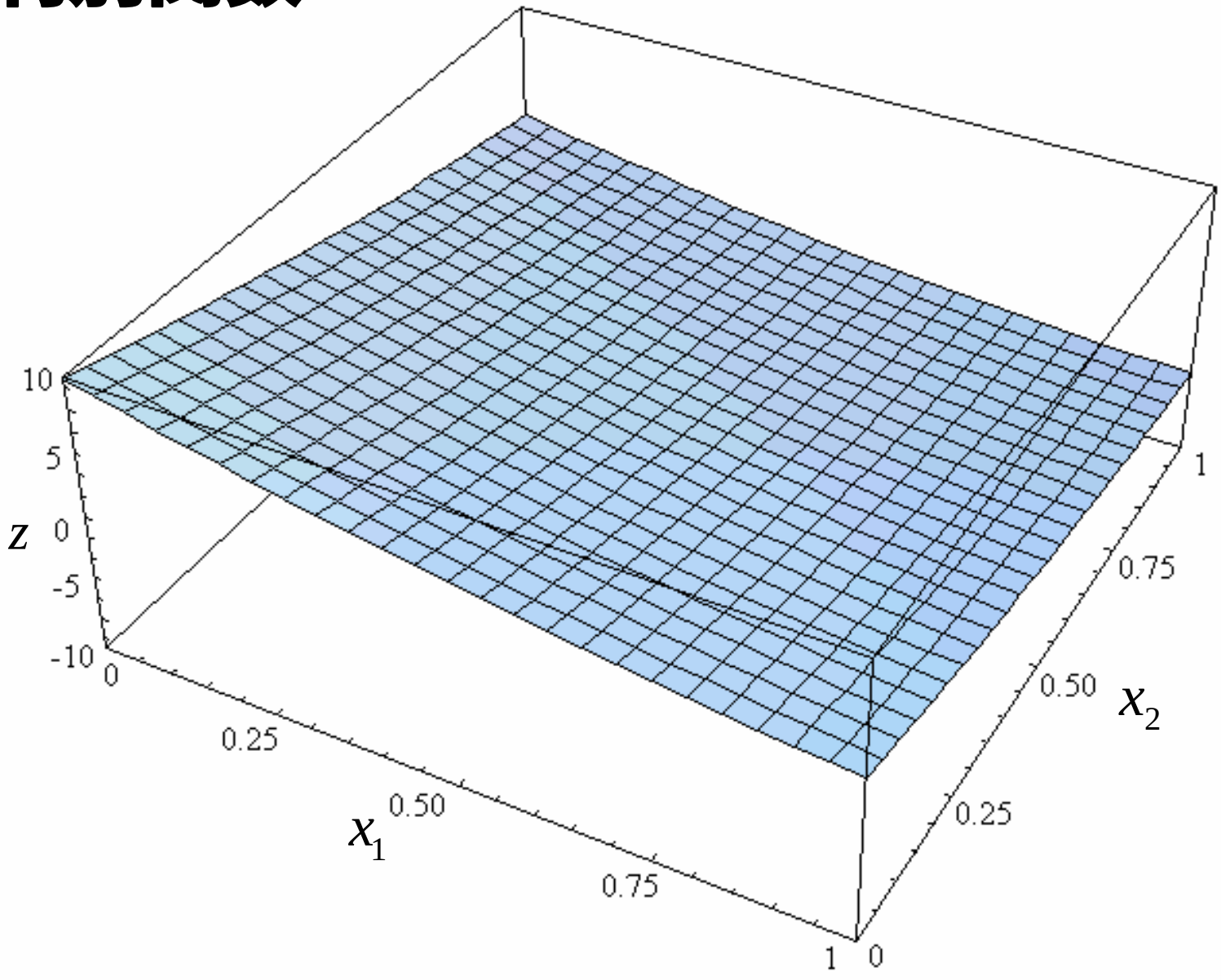
	誤判別率 (訓練)	誤判別率 (検証)
2次判別	0.221	0.259

線形判別関数　（$z<0$ なら第1群, $z>0$ なら第2群）



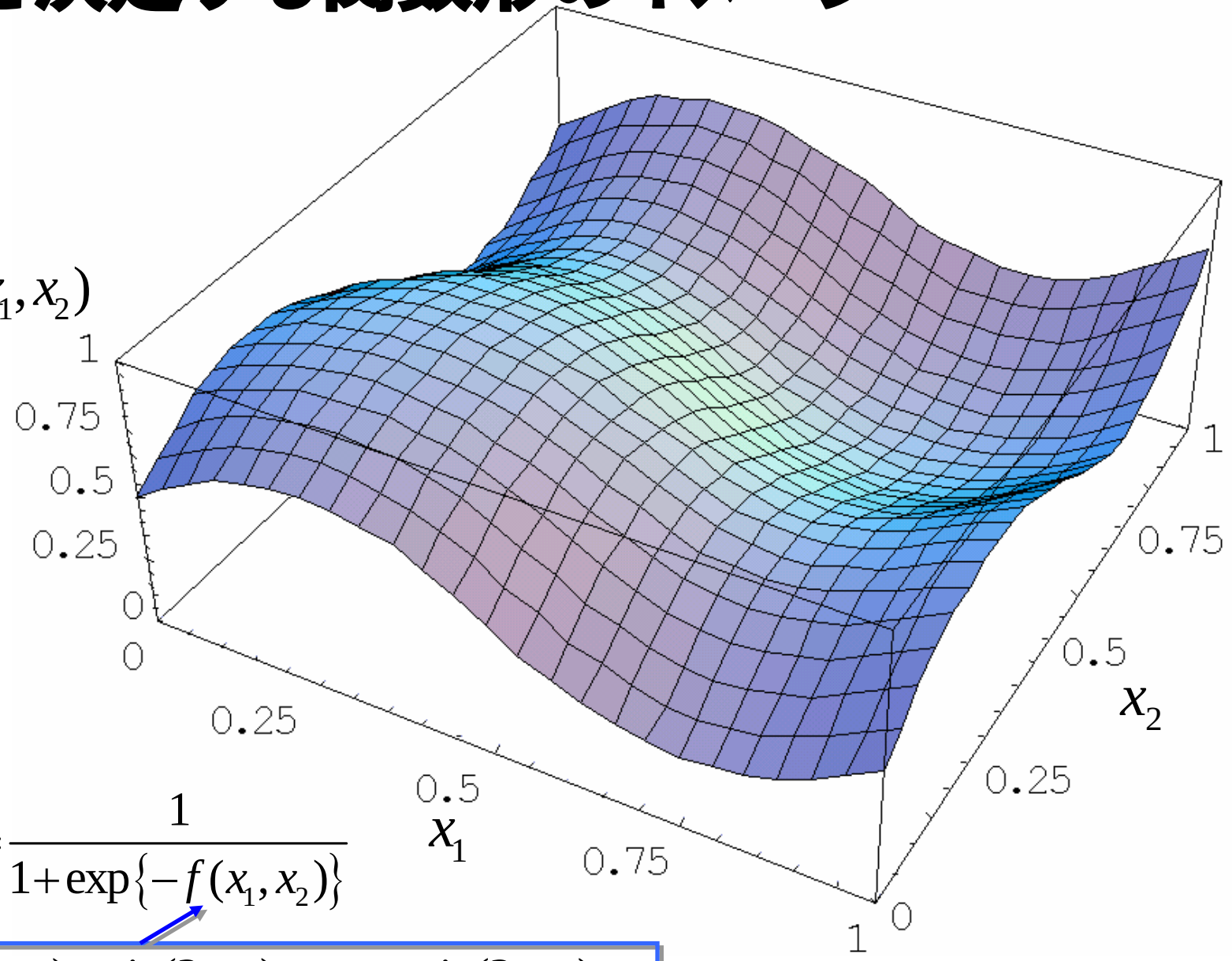
$$z = 4.33 - 3.82x_1 - 4.32x_2$$

2次判別関数　（$z<0$ なら第1群, $z>0$ なら第2群）



$$z = 12.85 - 15.38x_1 - 14.95x_2 + 5.04x_1^2 + 3.26x_1x_2 + 3.90x_2^2$$

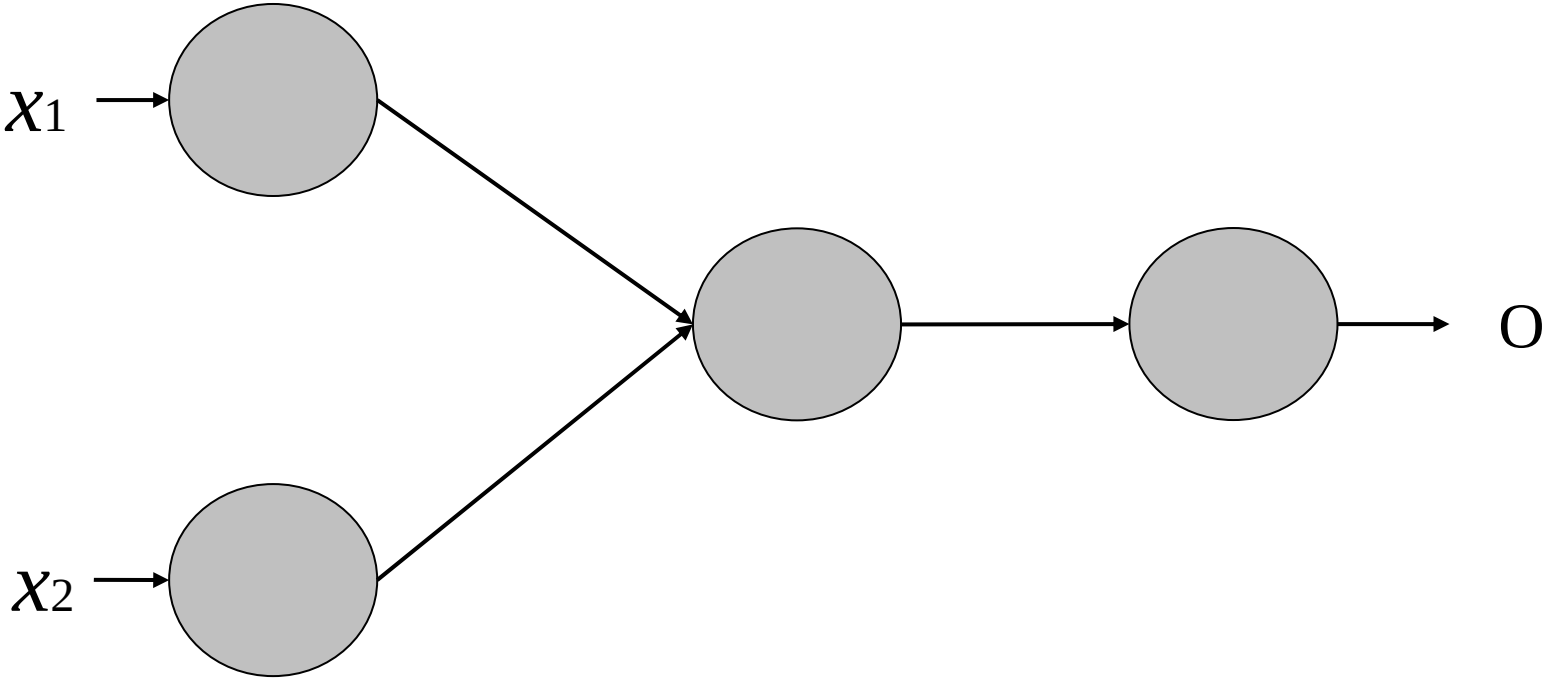
群を決定する関数形のイメージ



$$P_r = \frac{1}{1+\exp\{-f(x_1, x_2)\}}$$

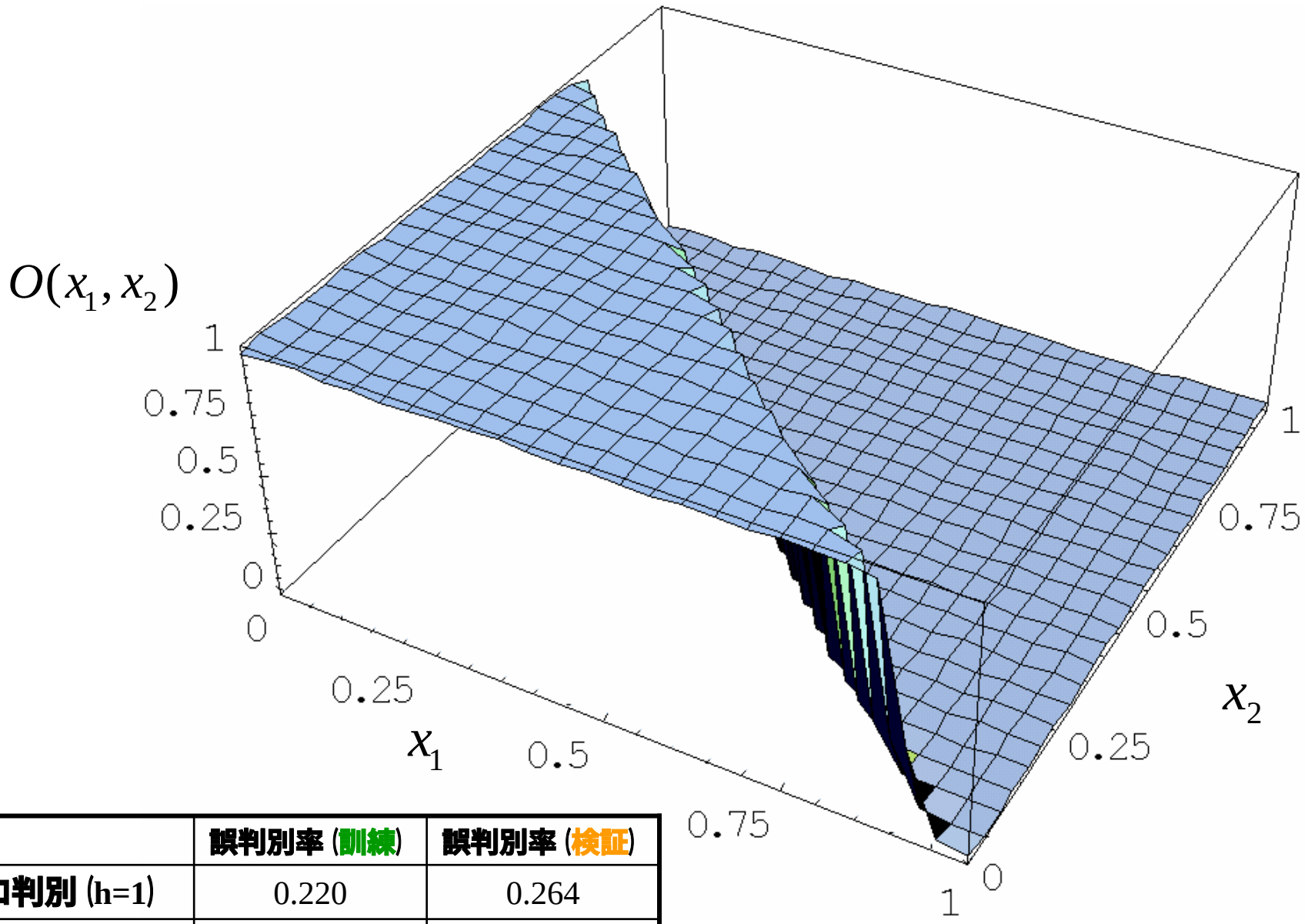
$$f(x_1, x_2) = \sin(2\pi x_1) + x_1 x_2 + \sin(2\pi x_2)$$

ロジスティック判別



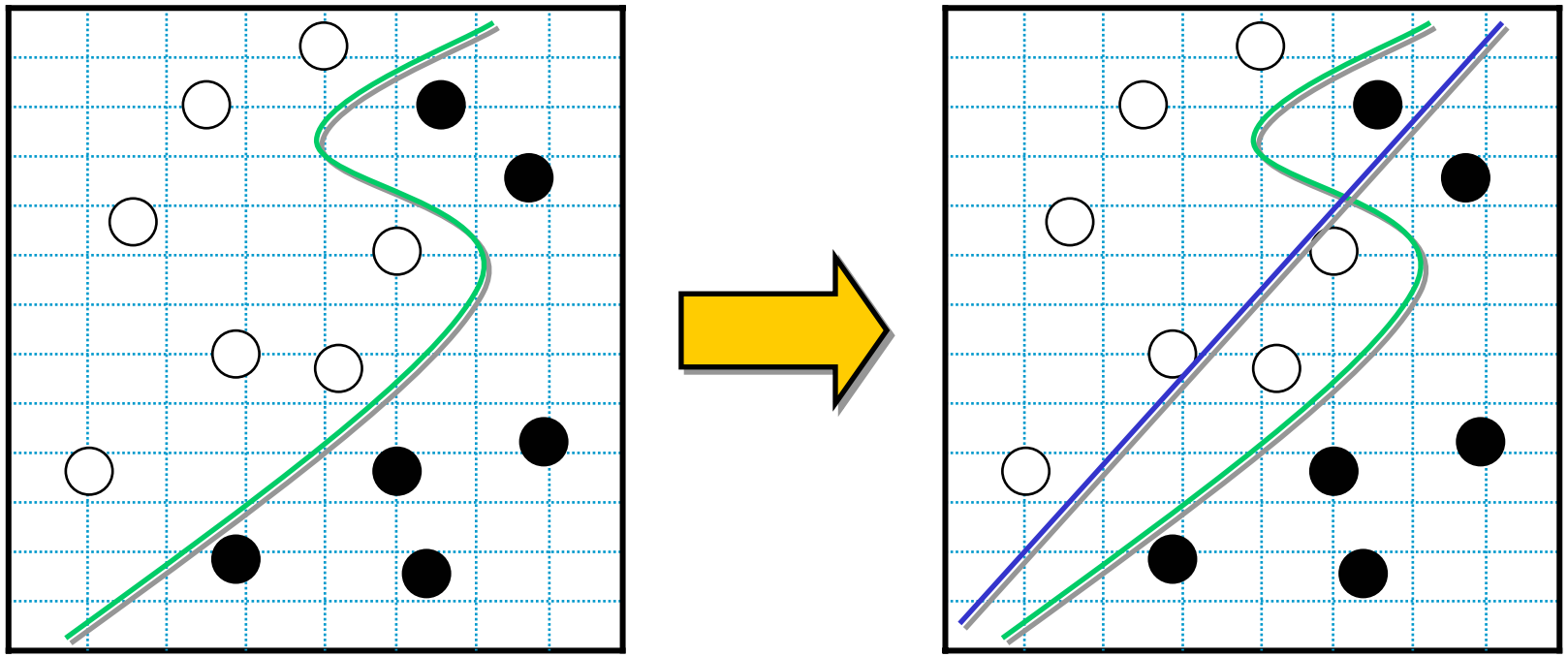
	誤判別率 (訓練)	誤判別率 (検証)
ロジスティック判別	0.220	0.262

NNによる近似（隠れユニット1）



	誤判別率（訓練）	誤判別率（検証）
ニューロ判別（h=1）	0.220	0.264
ロジスティック判別	0.220	0.262

本研究のシミュレーションは非線形



線形・準線形モデルの当てはまりが悪いのは当たりまえ.

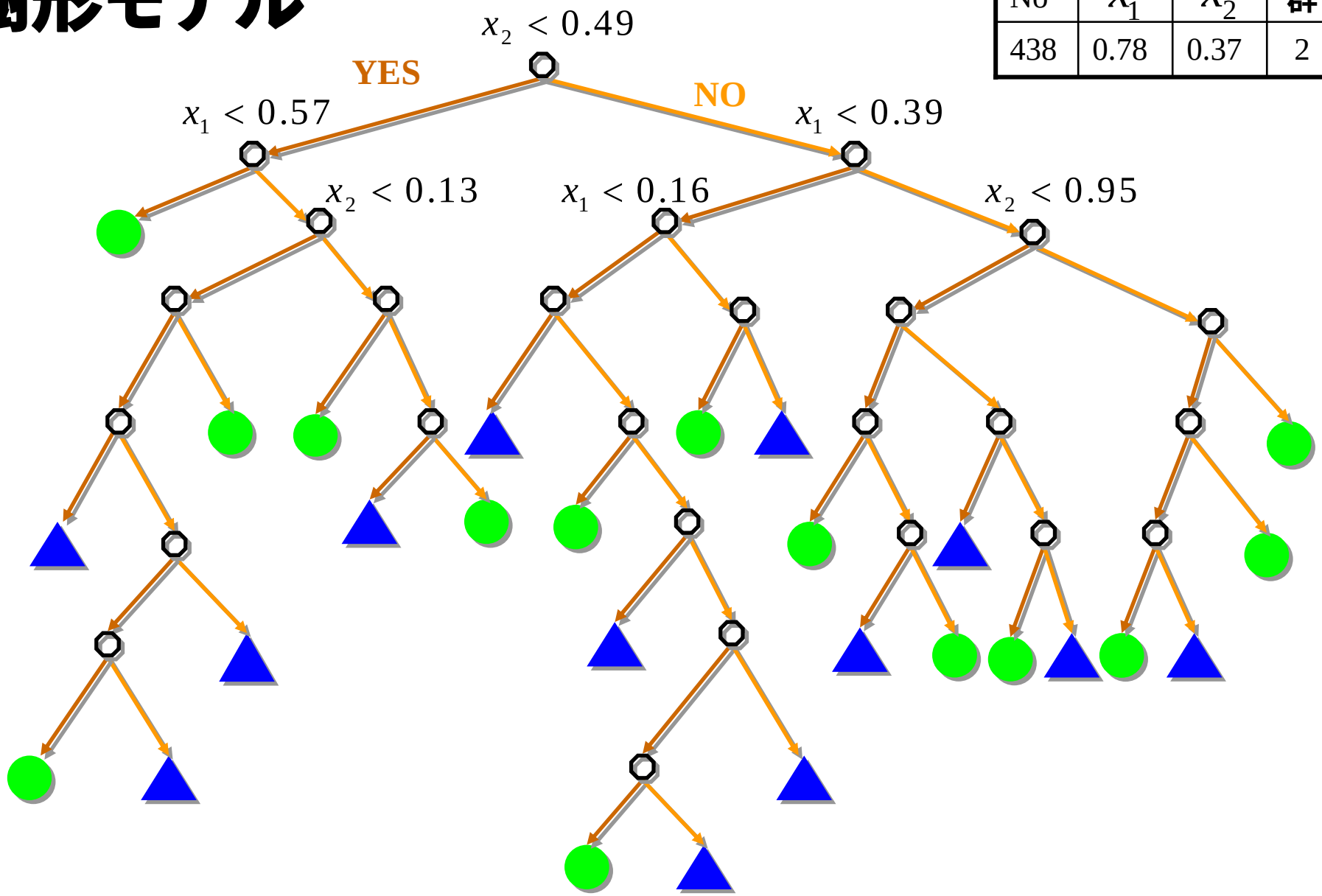
⇒他の非線形モデルとの性能比較.

他のモデルとの性能比較　その2

	誤判別率 (訓練)	誤判別率 (検証)
ニューロ判別モデル (隠れユニット数4)	0.003	0.012
線形判別 (2次判別)	0.234 (0.221)	0.266 (0.259)
ロジスティック判別	0.220	0.262
樹形モデル (CART)	?	?
Support Vector Machine	?	?
平滑化スプライン	?	?

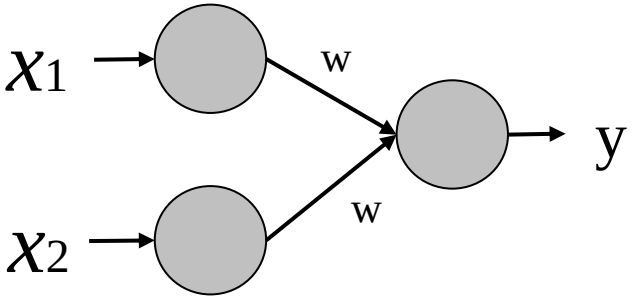
樹形モデル

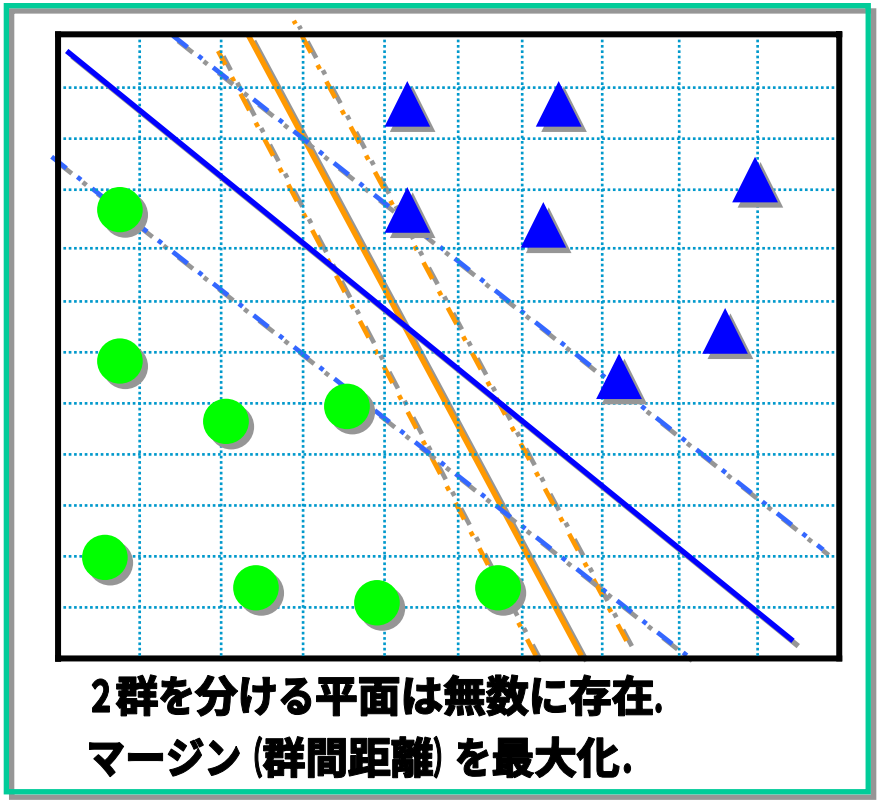
No	x_1	x_2	群
438	0.78	0.37	2



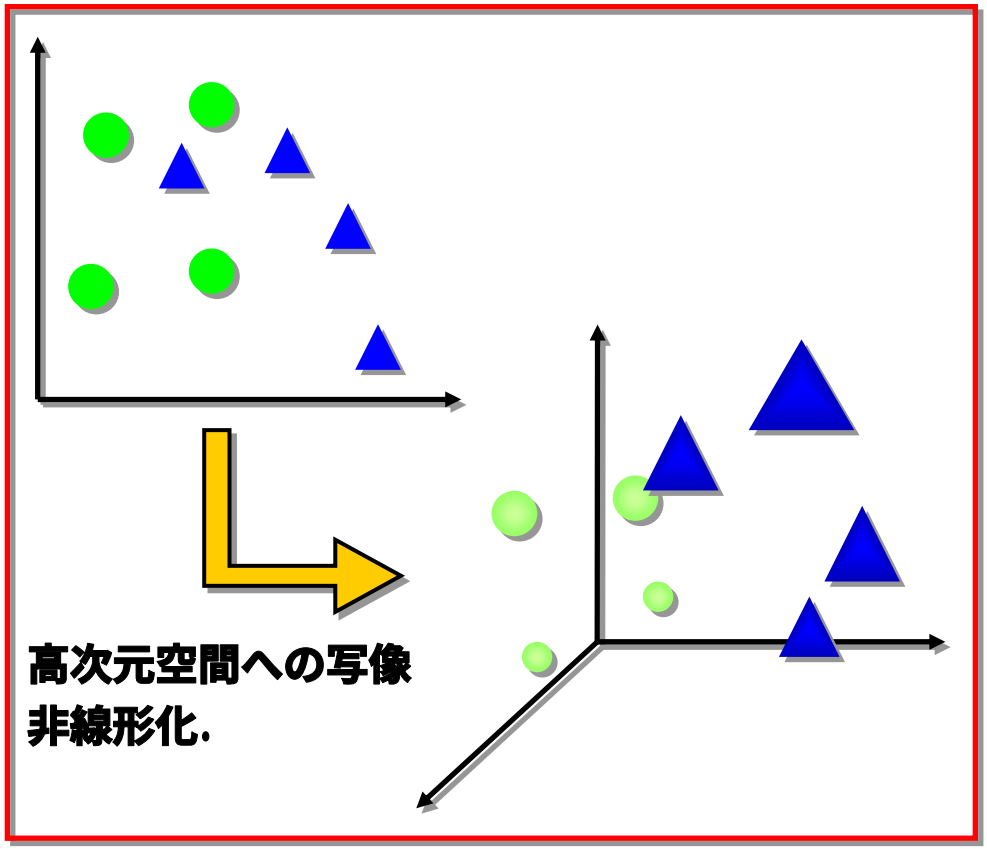
	誤判別率 (訓練)	誤判別率 (検証)
樹形モデル	0.023	0.066

Support Vector Machine





2群を分ける平面は無数に存在.
マージン (群間距離) を最大化.

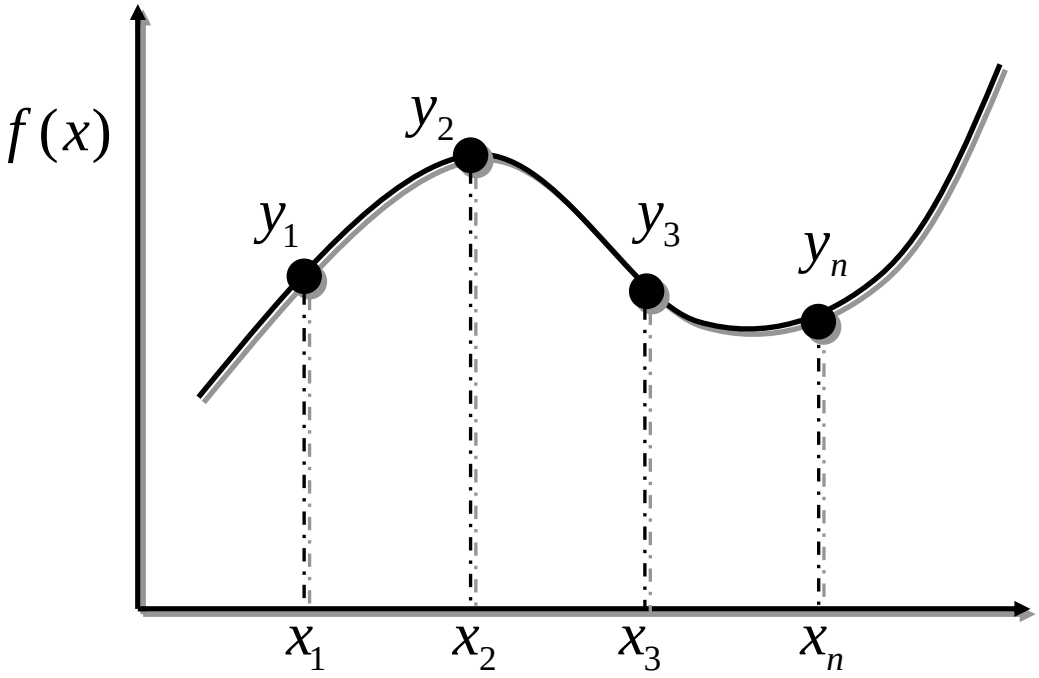


高次元空間への写像
非線形化.

	誤判別率 (訓練)	誤判別率 (検証)
SVM	0.053	0.072

平滑化スプライン

自然スプラインを利用し,サンプルから与えられた値より滑らかな曲線を得ることができる.

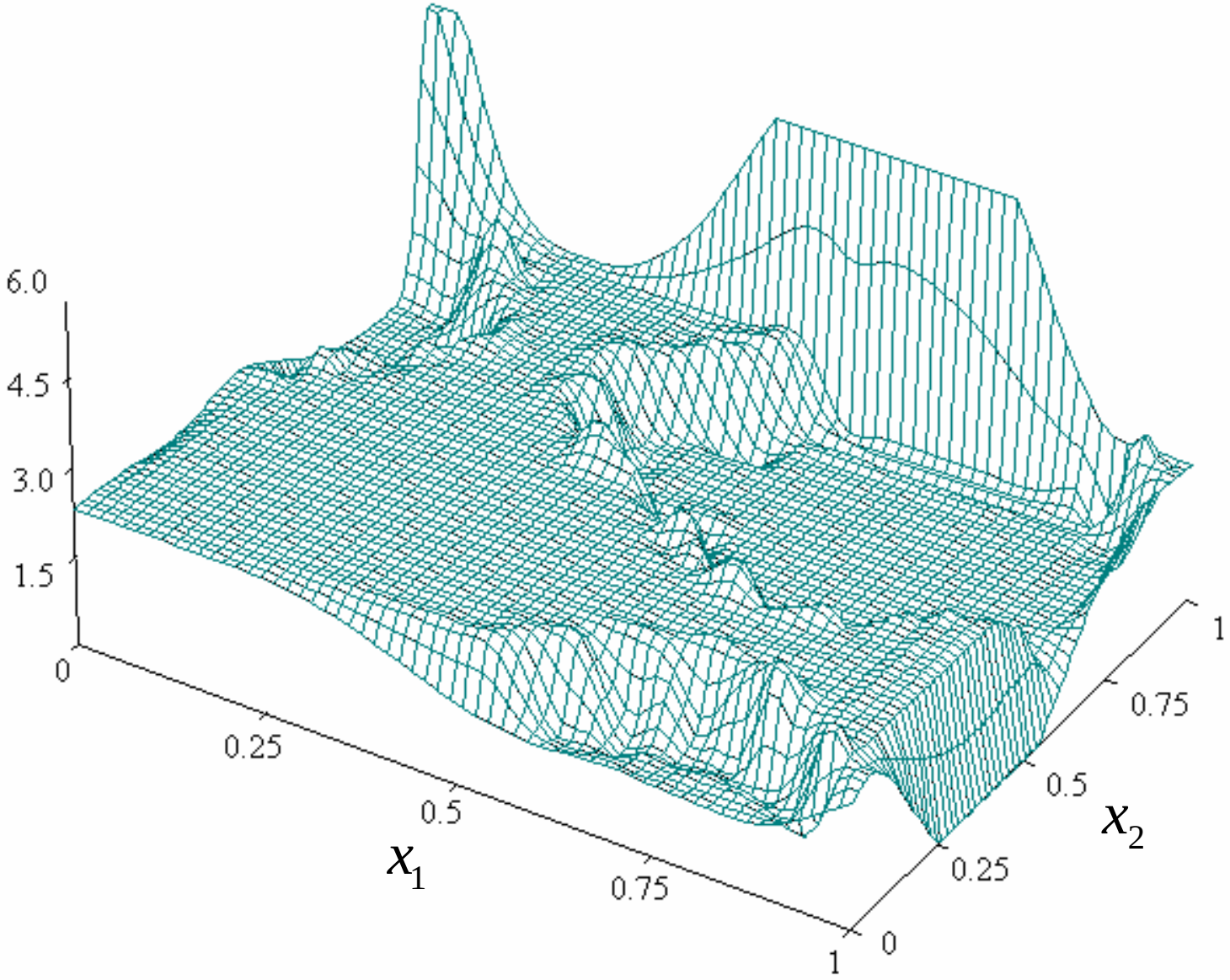


[3次の自然スプラインの例]

2項分布を仮定して,スプライン曲線を当てはめる.

	誤判別率 (訓練)	誤判別率 (検証)
平滑化スプライン	0.050	0.049

平滑化スプラインによる関数の近似



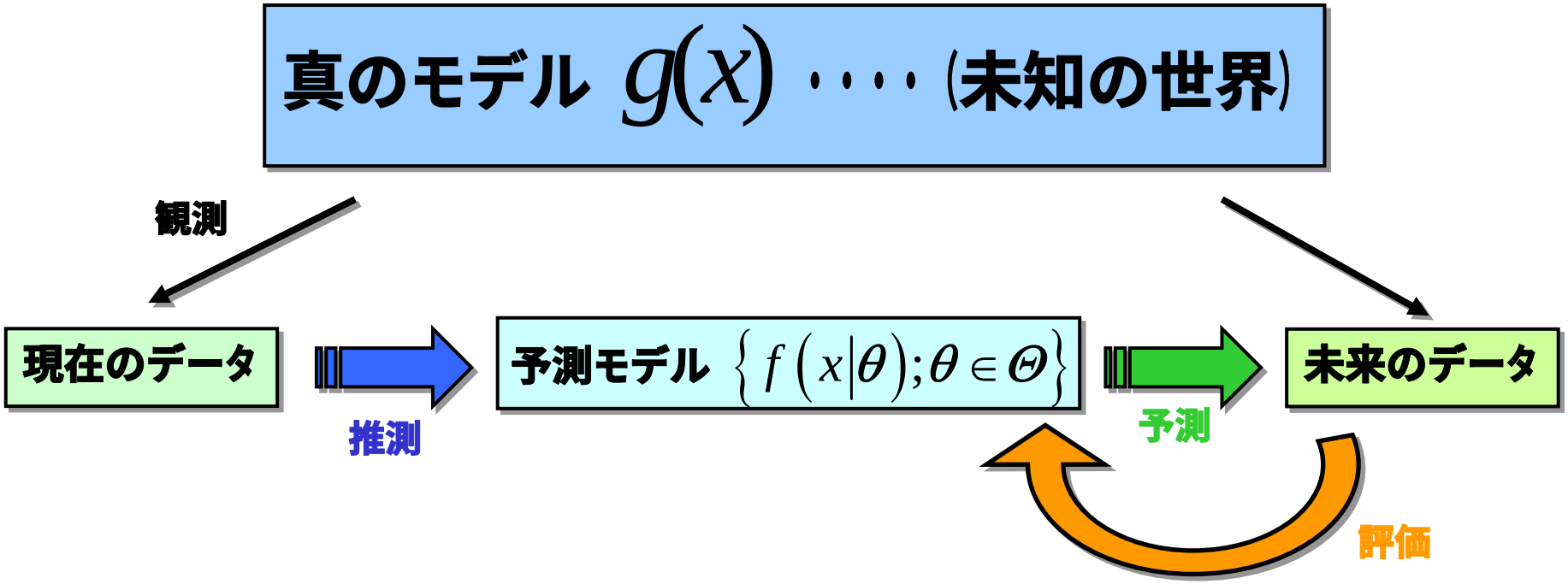
	誤判別率 (訓練)	誤判別率 (検証)
平滑化スプライン	0.050	0.049

性能比較の結果

	誤判別率 (訓練)	誤判別率 (検証)
ニューロ判別モデル (隠れユニット数4)	0.003	0.012
線形判別 (2次判別)	0.234 (0.221)	0.266 (0.259)
ロジスティック判別	0.220	0.262
樹形モデル (CART)	0.023	0.066
Support Vector Machine	0.053	0.072
平滑化スプライン	0.050	0.049

5. 今回のまとめ

- シミレーションデータを用いて近似精度を比較
- 検証標本を用いてモデルの検証
- 他のモデルとの性能比較


真のモデル $g(x)$ ・・・・(未知の世界)
観測
現在のデータ
推測
予測モデル $\{f(x|\theta);\theta\in\Theta\}$
予測
未来のデータ
評価