

ANFIS

Adaptive-Network-based Fuzzy Inference
System

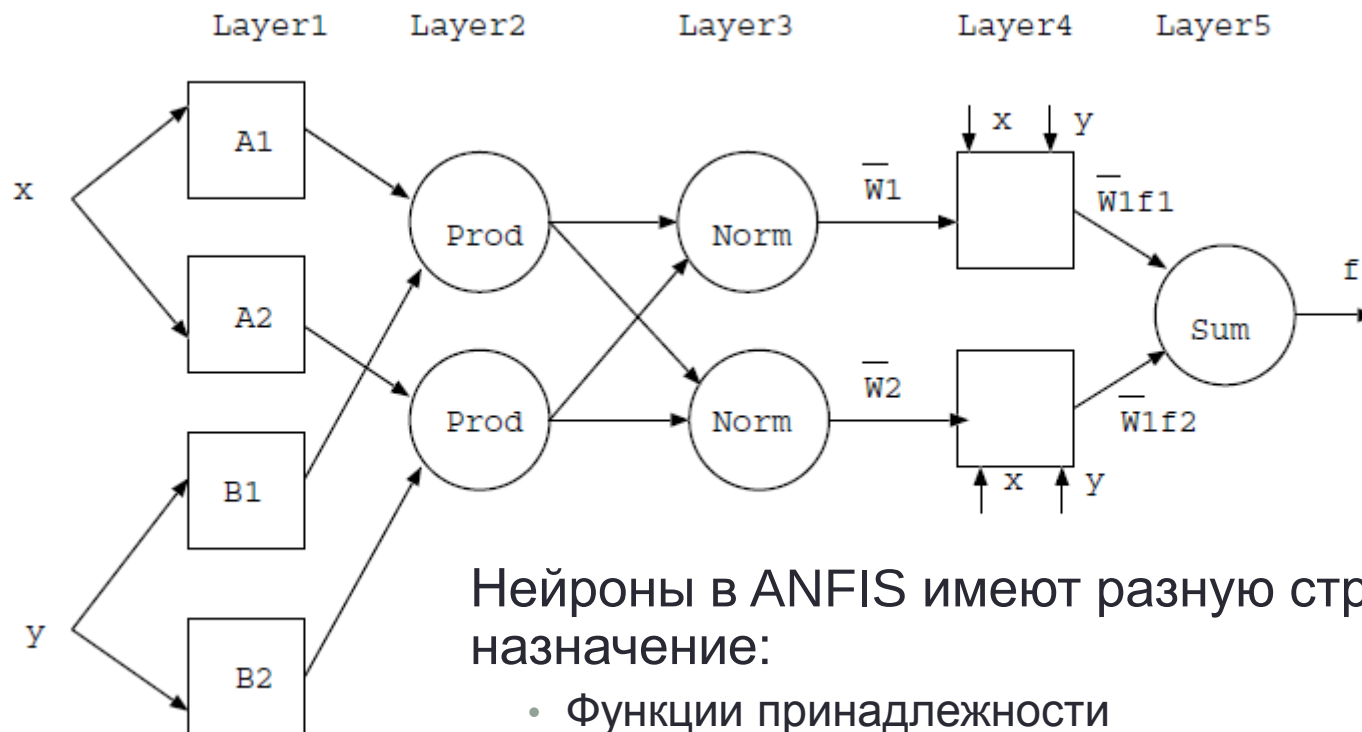
Введение

- ANFIS – адаптивная сеть, функционально эквивалентная системе нечёткого вывода.
- ANFIS использует гибридный алгоритм обучения.
- Алгоритм предложен J.-S. Roger Yang в 1992 году.
- ANFIS обычно использует нечёткий вывод Сугено.
 - Пусть имеется система нечёткого вывода с двумя входами, x и y и одним выходом z .
 - Модель нечёткого вывода Сугено 1-го порядка будет использовать правила вида

$$\textit{If } x \textit{ is } A_1 \textit{ and } y \textit{ is } B_1 \textit{ then } z = f_1 = p_1x + q_1y + r_1.$$

$$\textit{If } x \textit{ is } A_2 \textit{ and } y \textit{ is } B_2 \textit{ then } z = f_2 = p_2x + q_2y + r_2.$$

Архитектура ANFIS

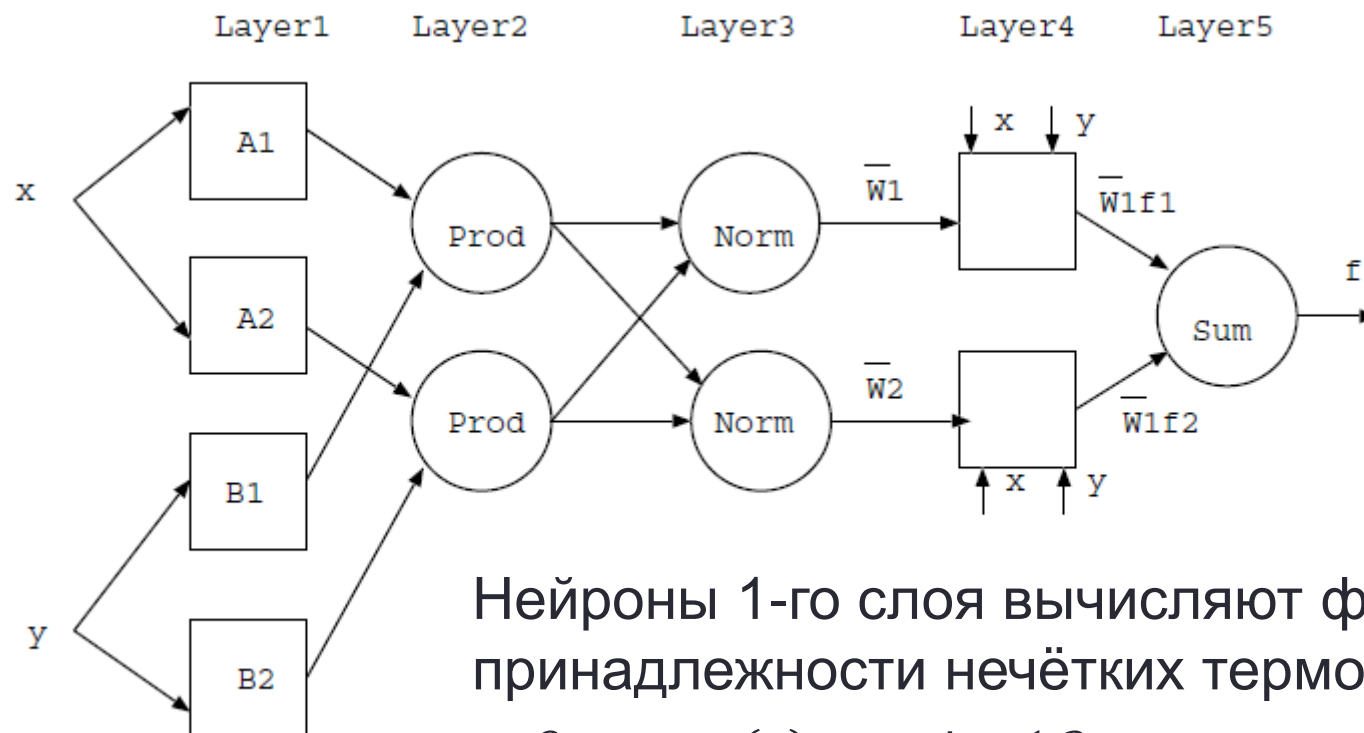


Нейроны в ANFIS имеют разную структуру и назначение:

- Функции принадлежности
- Правила
- Нормализация
- Линейные функции
- Выходное значение (сумма)

Обозначим $o_{l,i}$ - выход i -го нейрона слоя l .

Архитектура ANFIS – слой 1



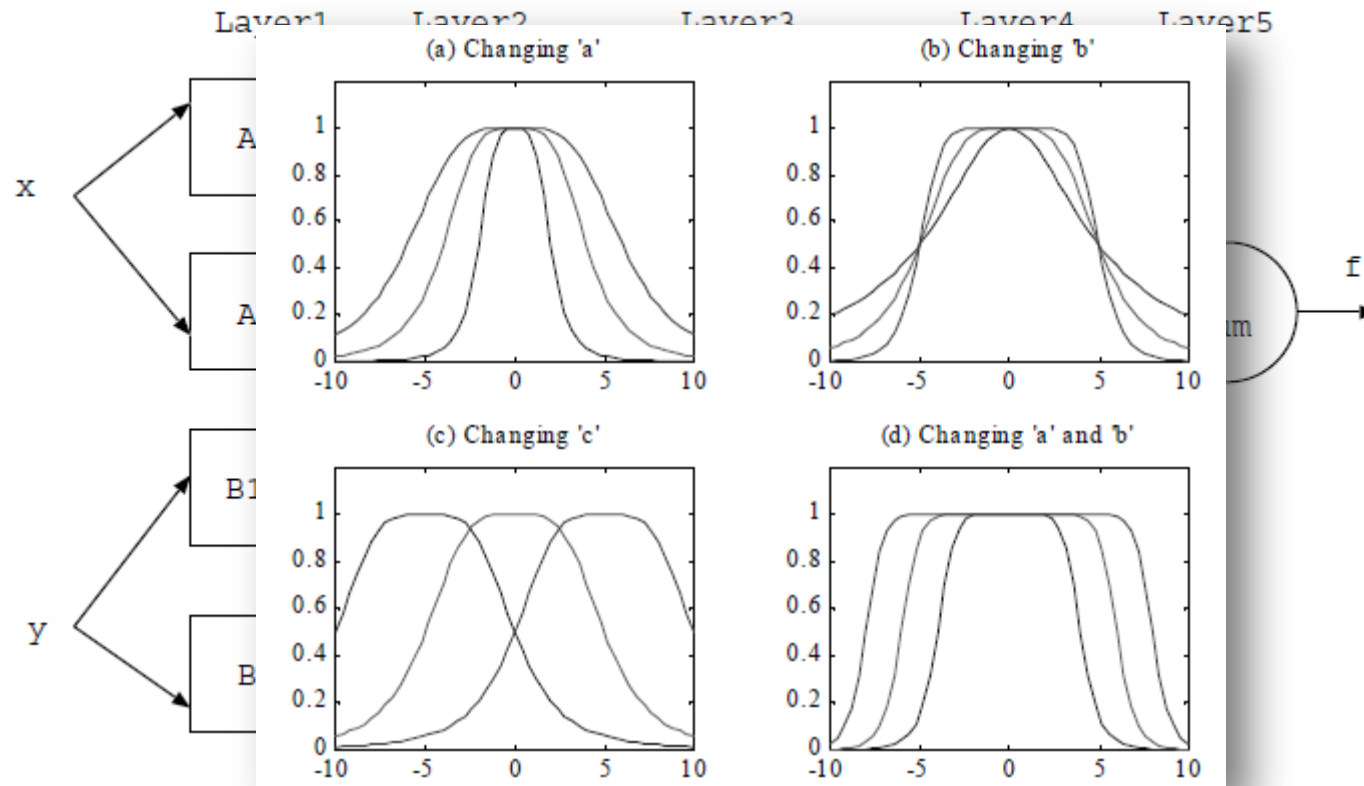
Нейроны 1-го слоя вычисляют функции принадлежности нечётких термов:

$$O_{1,i} = \mu_{A_i}(x), \text{ для } i = 1, 2,$$

$$O_{1,i} = \mu_{B_{i-2}}(x), \text{ для } i = 3, 4.$$

Слой является адаптивным.

Архитектура ANFIS – слой 1

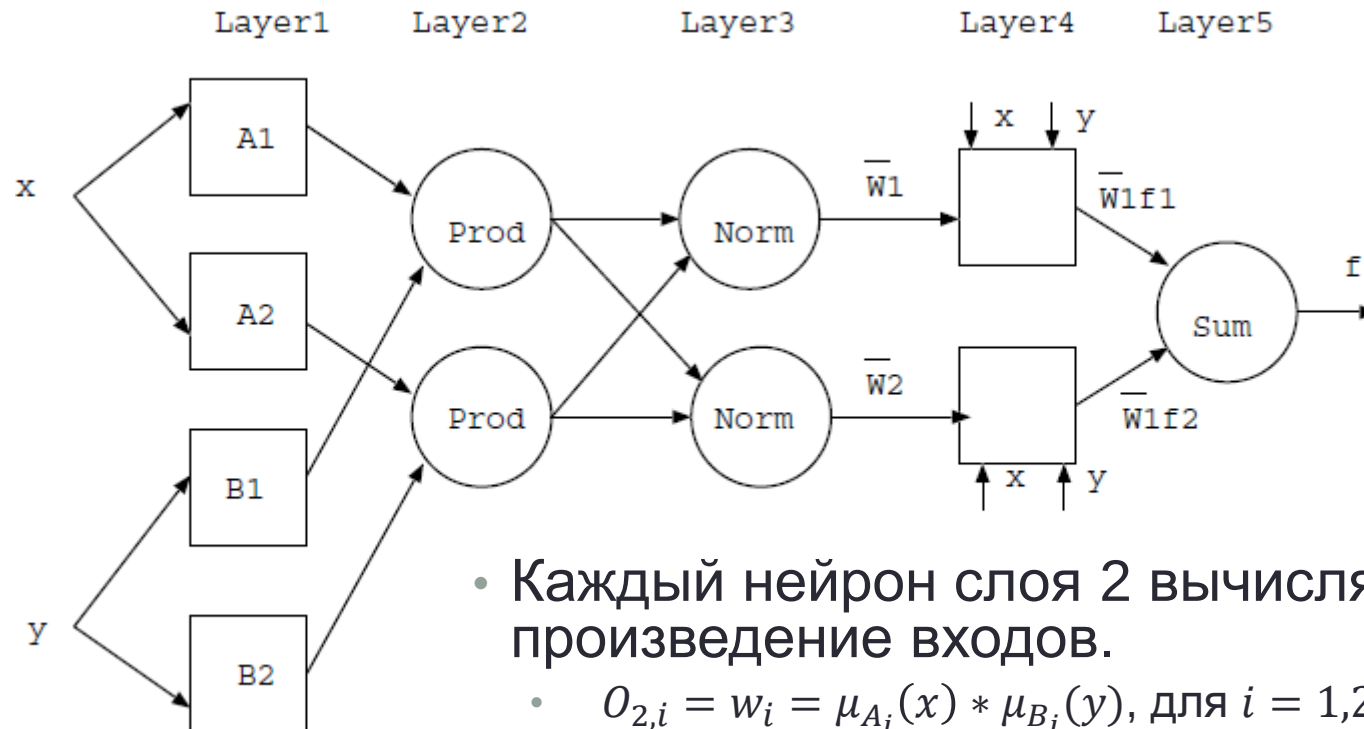


Типичная функция принадлежности:

$$\mu(x) = \frac{1}{1 + \left| \frac{x - c}{a} \right|^{2b}},$$

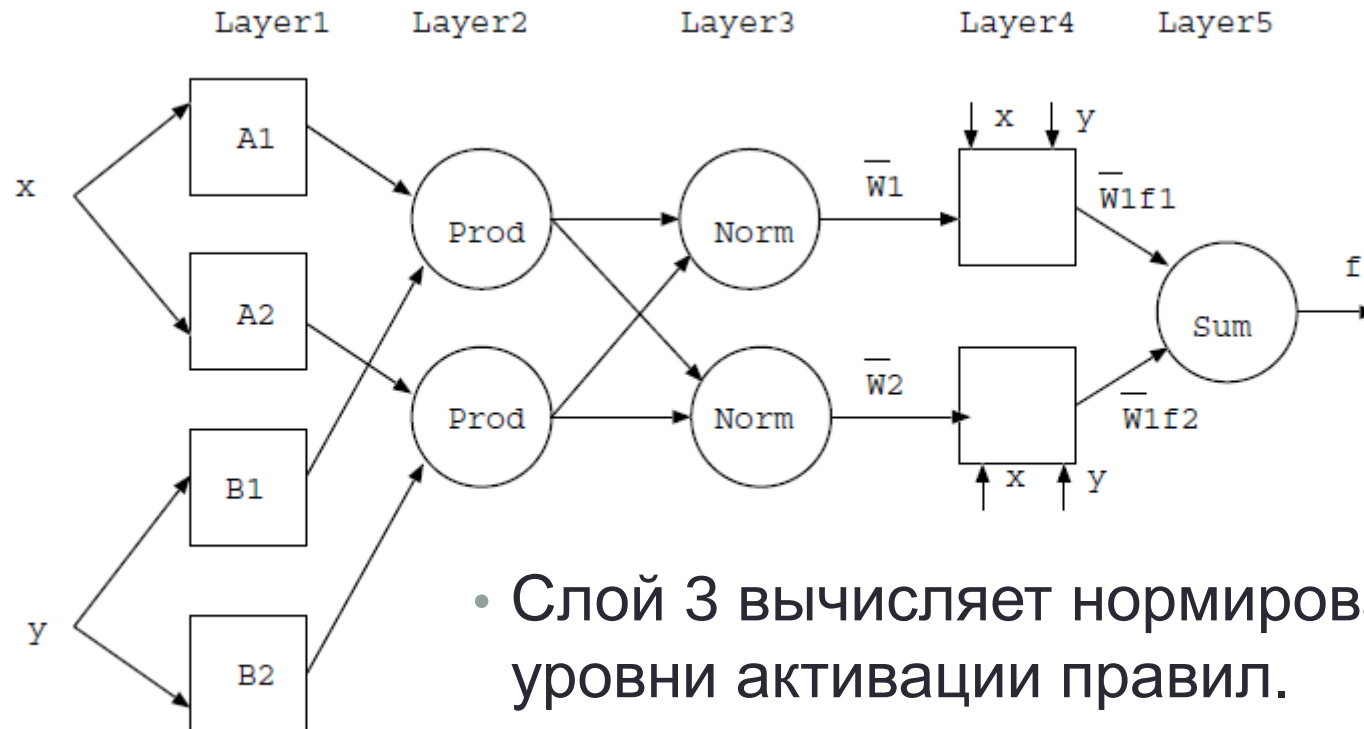
где a, b, c – параметры.

Архитектура ANFIS – слой 2



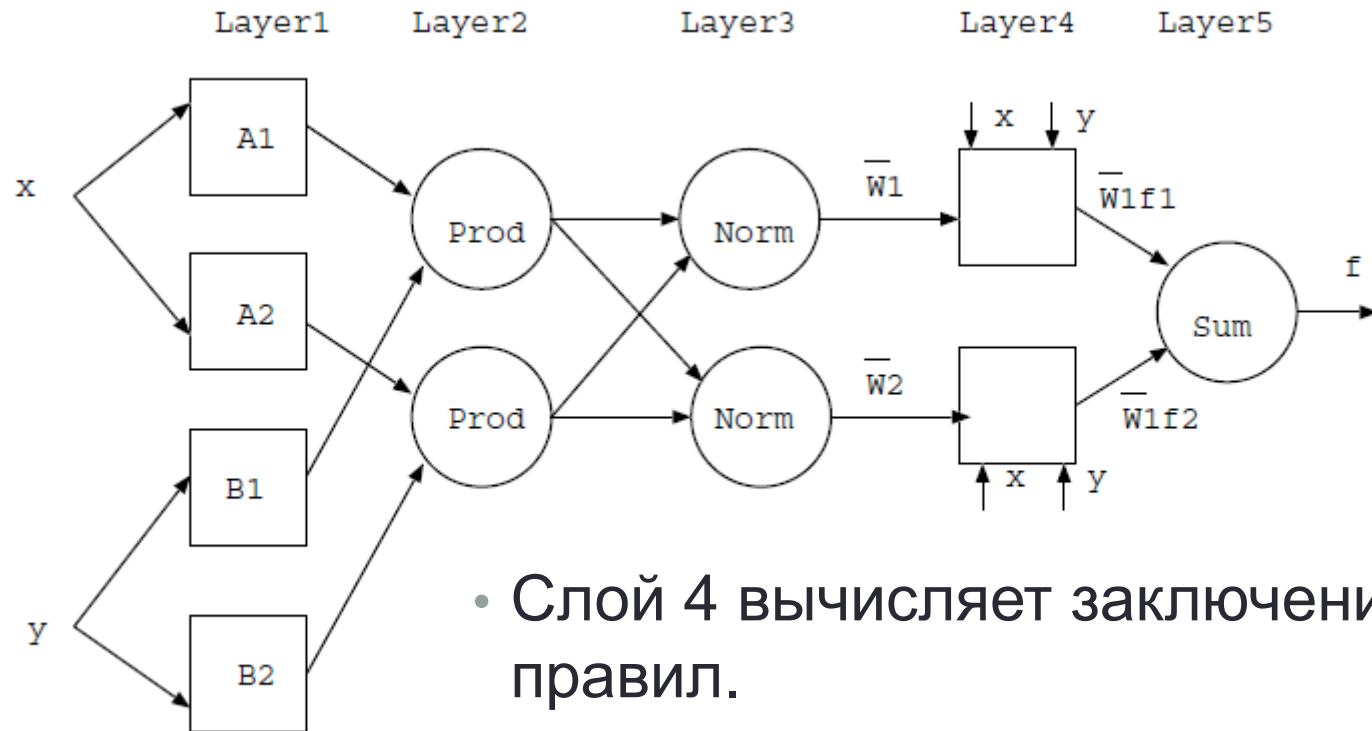
- Каждый нейрон слоя 2 вычисляет произведение входов.
 - $O_{2,i} = w_i = \mu_{A_i}(x) * \mu_{B_i}(y)$, для $i = 1, 2$,
- Выход нейрона представляет уровень активации правила.
- Могут использоваться другие t-нормы.
- Слой является фиксированным.

Архитектура ANFIS – слой 3



- Слой 3 вычисляет нормированные уровни активации правил.
 - $O_{3,i} = \bar{w}_i = \frac{w_i}{w_1 + w_2}$, для $i = 1, 2$.
- Слой является фиксированным.

Архитектура ANFIS – слой 4

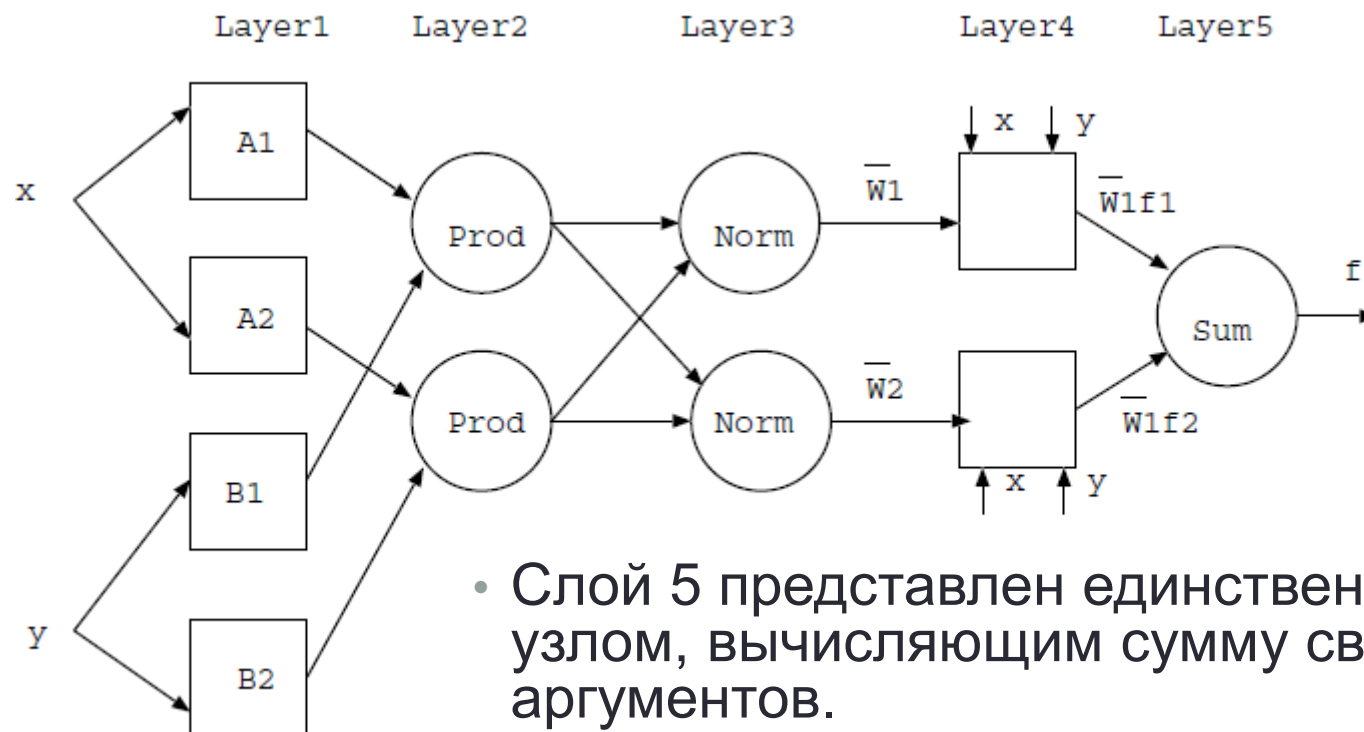


- Слой 4 вычисляет заключения правил.

$O_{4,i} = \bar{w}_i f_i = \bar{w}_i (p_i x + q_i y + r_i)$, для $i = 1, 2$,
где p_i, q_i, r_i - параметры узла.

- Слой является адаптивным.

Архитектура ANFIS – слой 5

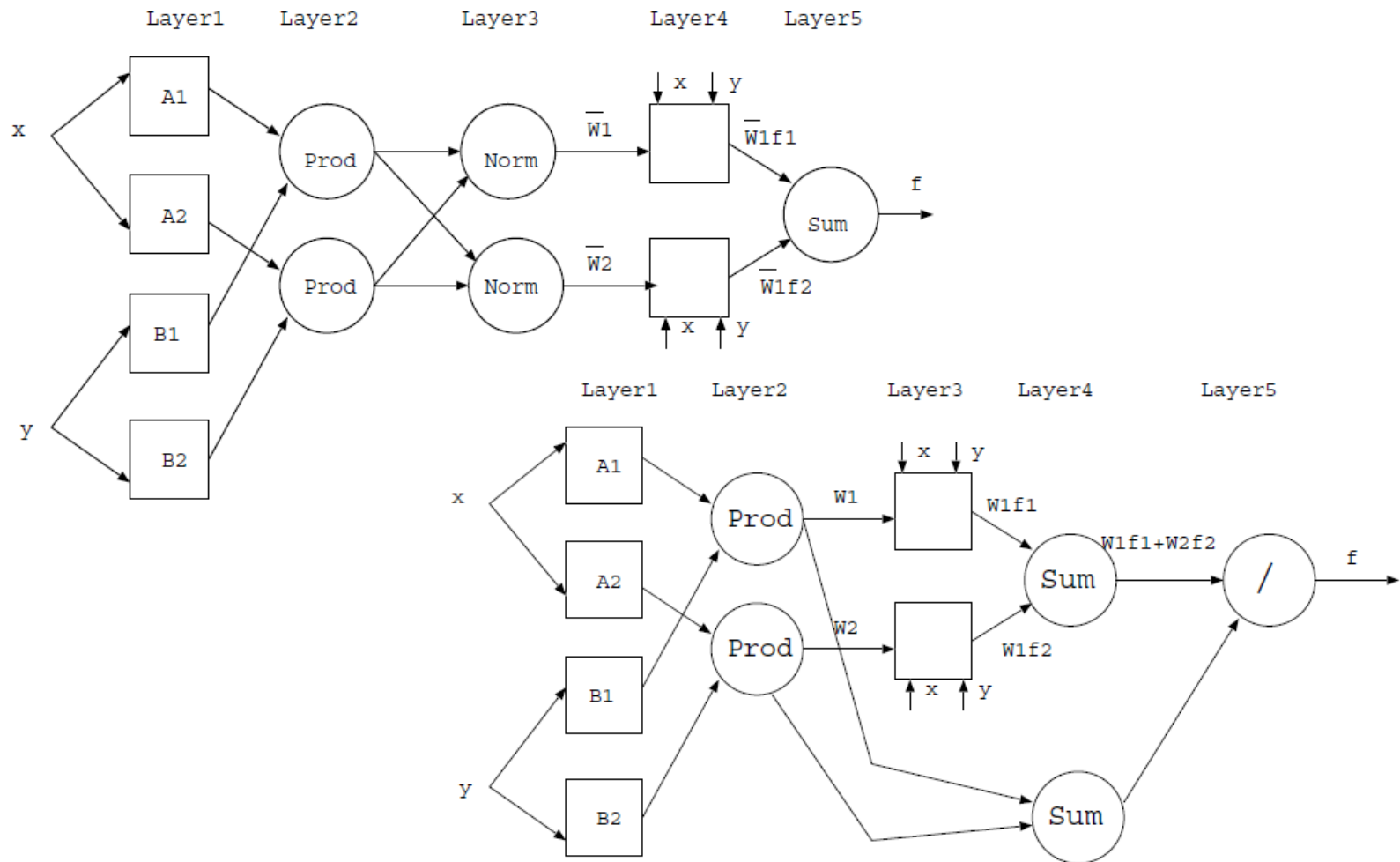


- Слой 5 представлен единственным узлом, вычисляющим сумму своих аргументов.

$$O_{5,1} = \sum_i \bar{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i}.$$

- Вычисляется результат нечёткого вывода.
- Слой является фиксированным.

Архитектура ANFIS – альтернативное представление



Алгоритм обучения

- Гибридный двухпроходный алгоритм:

прямой
ход

- Определяем параметры слоя 4.
- Используется метод наименьших квадратов.

обратный
ход

- Уточняем параметры первого слоя.
- Используется обратное распространение ошибки.

прямой
ход

обратный
ход

	Прямой ход	Обратный ход
Параметры предпосылок	Фиксированы	Градиентный спуск
Параметры заключений	МНК	Фиксированы

Алгоритм обучения - прямой ход

- Пусть $S_1 = \{\{a_{11}, b_{11}, c_{11}\}, \dots, \{a_{np}, b_{np}, c_{np}\}\}$ – параметры предпосылок, $S_2 = \{\{q_{10}, q_{11}, \dots, q_{1n}\}, \dots, \{q_{p^{n_0}}, \dots, q_{p^{n_n}}\}\}$ – параметры заключений, $S = S_1 \cup S_2$.
- При фиксированных параметрах из S_1 имея K обучающих примеров, можно представить задачу обучения в виде $AX = B$, где X содержит параметры из S_2 .
- Так как число обучающих примеров обычно больше числа параметров в S_2 , то точное решение данной системы уравнений не возможно, но можно воспользоваться методом наименьших квадратов для нахождения аппроксимации X^* , которая минимизирует ошибку $\|AX - B\|^2$:

$$X^* = (A^T A)^{-1} A^T B.$$

Алгоритм обучения - прямой ход

- Для вычислений МНК используется итеративная процедура:

$$S_{i+1} = S_i - \frac{S_i a_{i+1} a_{i+1}^T S_i}{1 + a_{i+1}^T S_i a_{i+1}},$$
$$X_{i+1} = X_i + S_{i+1} a_{i+1} (b_{i+1}^T - a_{i+1}^T X_i),$$

где:

- $X_0 = 0$;
- $S_0 = \gamma I$, γ - большое число, I – единичная матрица;
- a_i^T - i -я строка матрицы A ;
- b_i^T - i -ый элемент вектора B ;
- $X^* = X_K$.

Алгоритм обучения – обратный ход

- Для каждого параметра α_i из S_1 обновление осуществляется по правилу:

$$\Delta\alpha_i = -\eta \frac{\partial E}{\partial \alpha_i},$$

где:

- $E = \sum_{j=1}^K (O_{5,1}(x^{(j)}) - z^{(j)})^2$ - ошибка на обучающей выборке;
- η – скорость обучения, $\eta = \frac{k}{\sqrt{\sum_i \frac{\partial E}{\partial \alpha_i}}}$;
- k – размер шага.

Преимущества и недостатки



Более быстрая сходимость, чем у обычных нейронных сетей.

Меньший объем обучающей выборки.

Компактная модель.

Автоматическая настройка параметров.

Гарантированная гладкость выходной функции.

Высокая пространственная сложность.

Знаки коэффициентов не всегда соответствуют знакам выхода.

Осцилляции при большом числе правил.

Нельзя использовать min или триангулярные функции.

Сильное влияние выбросов.

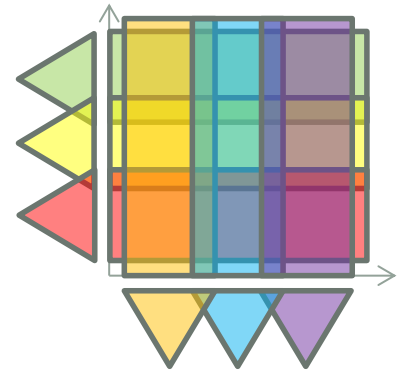
Симметричная функция ошибки.

ПОСТРОЕНИЕ СИСТЕМ НЕЧЁТКОГО ВЫВОДА С МНОГОМЕРНЫМИ ТЕРМАМИ

Нечёткий вывод с многомерными термами

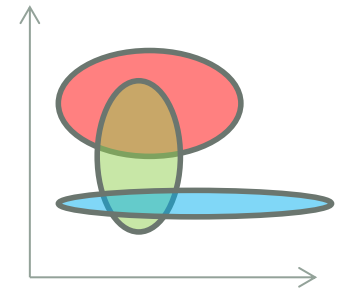
- В рассмотренных ранее системах
 - Термы предпосылок определялись отдельно для каждой входной переменной.
 - Множество правил образуется декартовым произведением терм-множеств разных переменных.
 - Правила имеют вид

If x_1 is A_1 and x_2 is B_1 then $z = \dots$



- Другая возможность
 - Термы являются многомерными нечёткими множествами, определёнными на всём входном пространстве.
 - Каждому терму соответствует одно правило.
 - Правила имеют вид

If x is A then $z = \dots$



Кластеризация и нечёткие системы

- Пусть имеется обучающая выборка $\{x^{(1)}, \dots, x^{(k)}\} = \{(y^{(1)}, z^{(1)}), \dots, (y^{(k)}, z^{(k)})\}$, где $y^{(i)}$ - вектор входных значений, $z^{(i)}$ - соответствующее выходное значение.
- Идея алгоритма (Stephen L. Chiu, 1994):
 1. Выполняем кластеризацию данных обучающей выборки (включая входную и выходную часть).
 2. Полученные кластеры рассматриваем как прототипы для нечётких правил:
 1. Пусть центр кластера имеет координаты (y^*, z^*) .
 2. В качестве координат центра соответствующей функции распределения для предпосылки правила можно взять y^* .
 3. Ширину функции распределения для предпосылки определим методом подбора.
 4. В качестве заключения правила для модели Сугено-0 в первом приближении возьмём z^* .
 3. Для уточнения заключения и получения параметров для модели Сугено-1 используем МНК.

Алгоритм кластеризации

Mountain Method

Yager, Filev, 1992:

1. Разбиваем пространство регулярной сеткой.
2. Для каждого узла сетки вычисляем потенциал, на основе расстояний до точек данных.
 - Чем больше рядом точек, тем больше потенциал.
3. Выбираем узел с максимальным потенциалом как центр кластера.
4. Уменьшаем потенциал остальных узлов пропорционально расстоянию до выбранного центра.
 - У соседних с центром узлов потенциал должен быть сильно уменьшен.
5. Если не выполнено условие остановки, перейти к 3.

Модифицированный алгоритм

- Пусть имеется n точек $\{x_1, \dots, x_n\}$, $x_i \in R^M$. Будем считать, что все координаты имеют одинаковый диапазон изменения (нормированы).
- В качестве потенциальных центров кластеров будем рассматривать сами точки x_i .
- Потенциал точки i определяется как

$$P_i = \sum_{j=1}^n e^{-\alpha \|x_i - x_j\|^2},$$

где $\alpha = 4/r_\alpha^2$, r_α - положительная константа, определяющая радиус соседства.

- Выберем в качестве центра кластера точку x^* с максимальным потенциалом. Обозначим этот потенциал P^* .
- Потенциал остальных точек уменьшим по правилу:

$$P_i = P_i - P^* e^{-\beta \|x_i - x^*\|^2},$$

где $\beta = 4/r_\beta^2$, r_β - положительная константа, например $r_\beta = 1.5r_\alpha$.

Критерий остановки

- Yager, Filev:

$P_k^* < \varepsilon P_1^*$, где P_i^* - потенциал i -го центра, ε – константа.

- Chiu:

if $P_k^* > \bar{\varepsilon} P_1^*$

Принять x_k^* как центр кластера и продолжить.

else if $P_k^* < \underline{\varepsilon} P_1^*$

Отвергнуть x_k^* как центр кластера и закончить работу.

else

$$d_{min} = \min_{i \in 1, \dots, k-1} \|x_k^* - x_i^*\|.$$

$$\text{if } \frac{d_{min}}{r_\alpha} + \frac{P_k^*}{P_1^*} \geq 1$$

Принять x_k^* как центр кластера и продолжить.

else

Отвергнуть x_k^* как центр кластера.

Установить потенциал точки x_k^* равным 0.

Проверить условие для следующего потенциального центра.

Здесь $\bar{\varepsilon}, \underline{\varepsilon}$ - константы. Авторы использовали $\bar{\varepsilon} = 0.5, \underline{\varepsilon} = 0.15$.

Построение правил

- Пусть центр i -го кластера находится в точке $x_i^* = (y_i^*, z_i^*)$.
- Функция принадлежности для предпосылки соответствующего правила

$$\mu_i(y) = e^{-\alpha \|y - y_i^*\|^2},$$

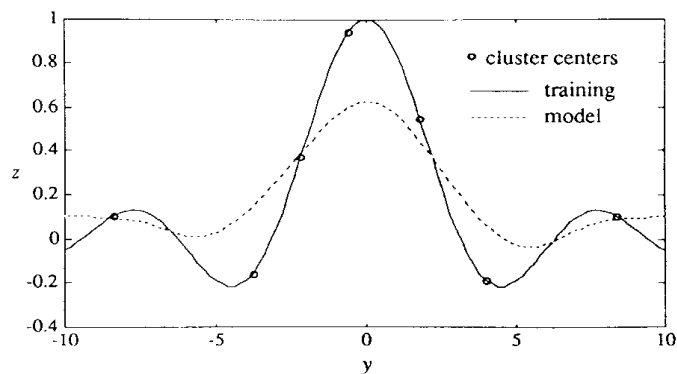
где $\alpha = 4/r_\alpha^2$, r_α - положительная константа из алгоритма кластеризации.

- Заключение правила для модели Сугено-0: $f = z_i^*$.
- Заключение могут быть оптимизированы с помощью МНК аналогично ANFIS.

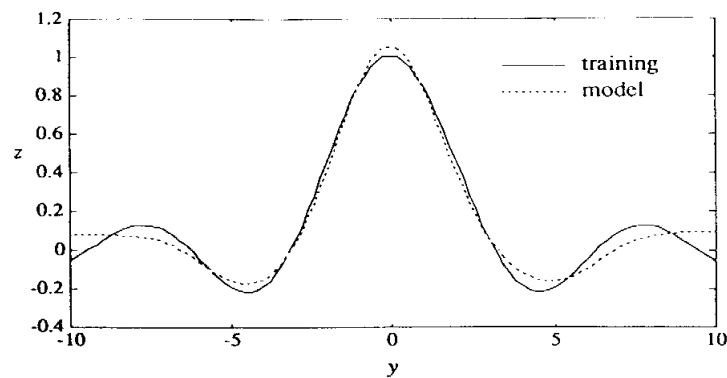
Пример

- Построение функции $z = \sin(y)/y$ на отрезке $[-10,10]$ по 100 точкам.

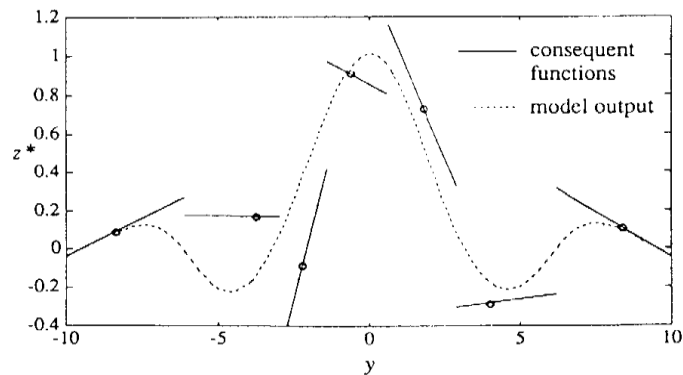
Сугено 0 без оптимизации



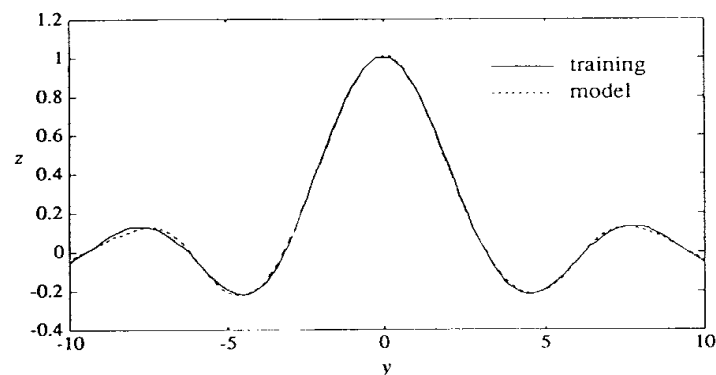
Сугено 0 с МНК



Правила Сугено 1



Сугено 1 с МНК



Оценка параметров многомерных возможностных распределений

Теорема 1. Пусть:

1. $f: R \rightarrow R$ — строго убывающая на луче $[0, +\infty)$ функция, $f(0) = 1$.
2. $Q = \{\mu_{\xi}(\mathbf{x}, \mathbf{c}, A)\}$ — семейство распределений, где $\mathbf{c} \in R^m$, A — положительно определённая симметричная матрица, $\mu_{\xi}(\mathbf{x}, \mathbf{c}, A) = \max \{0, f(\sqrt{(\mathbf{x} - \mathbf{c})^T A (\mathbf{x} - \mathbf{c})})\}$.
3. $X = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$, где $\mathbf{x}_i \in R^m, i \in [1, \dots, n]$ — наблюдаемые значения возможностной величины, принадлежащей семейству распределений Q .
4. W — многомерный эллипсоид минимального объёма, описанный вокруг множества X .

Тогда максиминной оценкой параметра $\theta = (\mathbf{c}, A)$ является $(\hat{\mathbf{c}}, \hat{A})$, такие, что эллипсоид W задаётся уравнением $(\mathbf{x} - \hat{\mathbf{c}})^T \hat{A} (\mathbf{x} - \hat{\mathbf{c}}) = 1/q^2$, $q = \arg \max_{t \geq 0} t^m \bar{f}(t)$.

Теорема: Оценка $\hat{\theta} = (\hat{\mathbf{c}}, \hat{A})$, полученная в теореме 1 является достаточной оценкой параметров $\mathbf{c}, A$ и состоятельной оценкой $\mathbf{c}$ при фиксированной матрице A .

Достаточная оценка представляет тот же объём информации о семействе, что и вся выборка.

Состоятельность отражает возможность того, что полученная оценка параметров представляет собой действительные значения параметров.

Пример восстановления параметров двумерного распределения

Рассмотрим μ/E оценку параметров двумерного распределения семейства $Q_\mu(\mathbf{c}, A)$

$$\text{при } f(t) = e^{-t^2}, q = \arg \max_{t \geq 0} t^2 \bar{f}(t) = 1$$

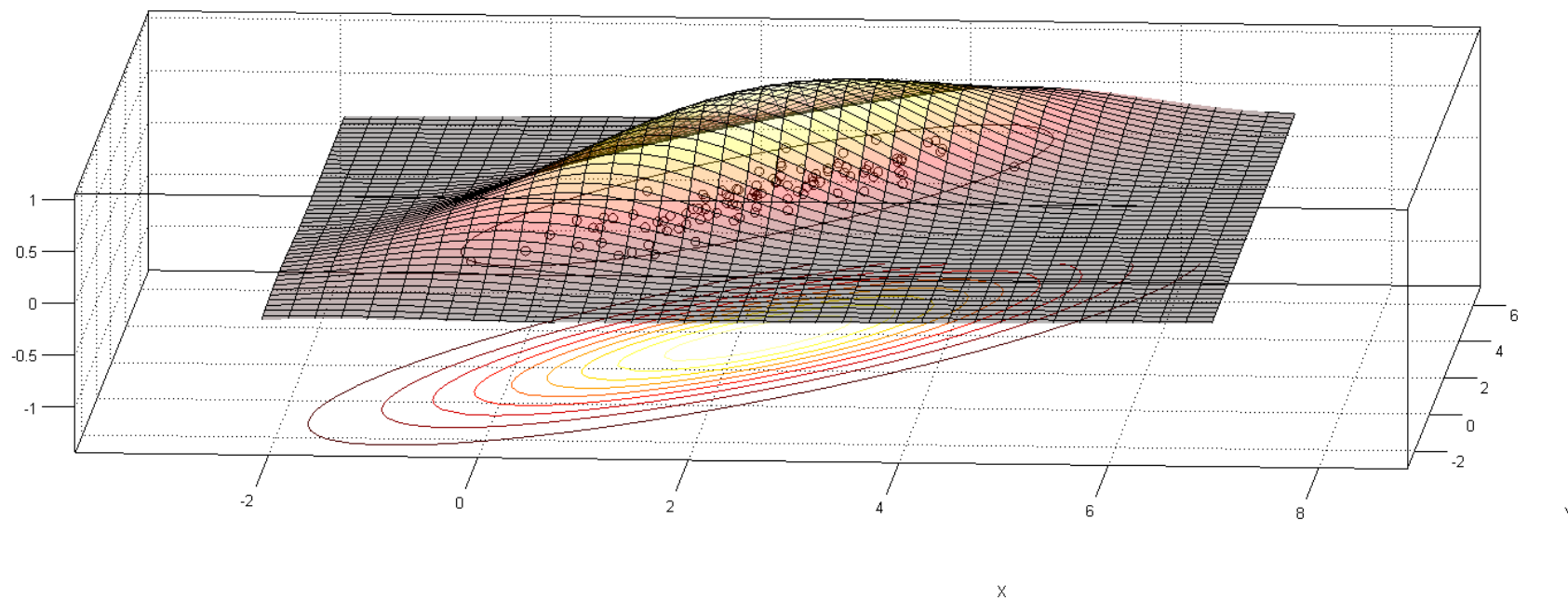


Рис. Восстановленное двумерное нормальное распределение

Использование μ/E оценки в построении систем нечёткого вывода 1

- Алгоритм 1:
 1. Кластеризовать обучающую выборку используя алгоритм k-средних.
 2. Для каждого кластера $C_i = \{x_{(i,1)}, \dots, x_{(i,n_i)}\}$ создать нечёткое правило, параметры предпосылки которого определяются с помощью μ/E оценки на множестве $\{y_{(i,1)}, \dots, y_{(i,n_i)}\}$.
 3. Определить заключения правил используя МНК.
- Наблюдения:
 - Итоговая ошибка аппроксимации зависит от результата работы алгоритма k-средних, который зависит от начального положения центров.
 - На простых примерах легко достигаются результаты, превосходящие результаты алгоритма Chiu.
 - На сложных примерах найти хорошую кластеризацию с помощью k-средних становится сложно, и алгоритм уступает алгоритму Chiu.

Использование μ/E оценки в построении систем нечёткого вывода 2

- Алгоритм 2:
 1. Кластеризовать обучающую выборку используя модифицированный Mountain Method.
 2. Отнести каждую точку к кластеру с ближайшим центром x_i^* .
 3. Для каждого кластера $C_i = \{x_{(i,1)}, \dots, x_{(i,n_i)}\}$ создать нечёткое правило, параметры предпосылки которого определяются с помощью μ/E оценки на множестве $\{y_{(i,1)}, \dots, y_{(i,n_i)}\}$.
 4. Определить заключения правил используя МНК.

Аппроксимация функции 5 переменных

Было осуществлено прогнозирование значений временного ряда, порождённого дифференциальным уравнением Mackey-Glass. Значение ряда прогнозируется по 5 предыдущим значениям.

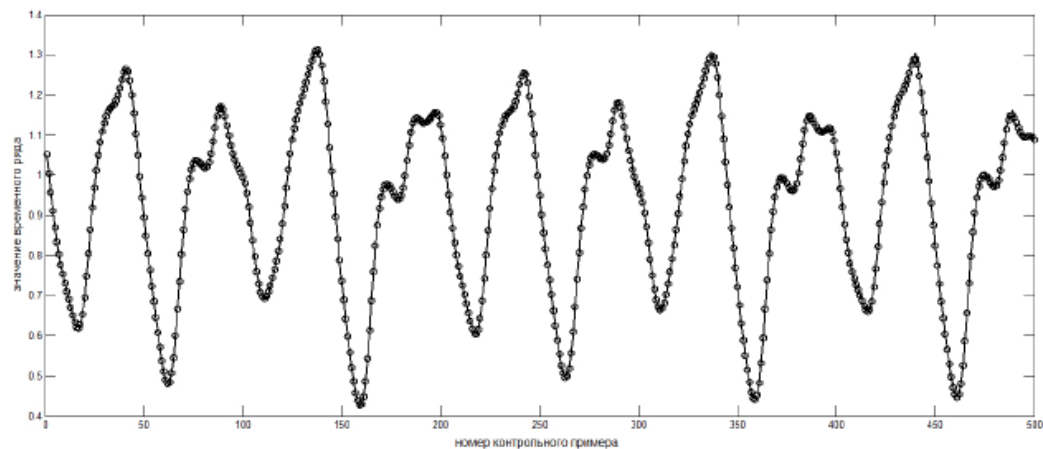


Рис. Прогноз значений временного ряда, порождённого уравнением Mackey-Glass

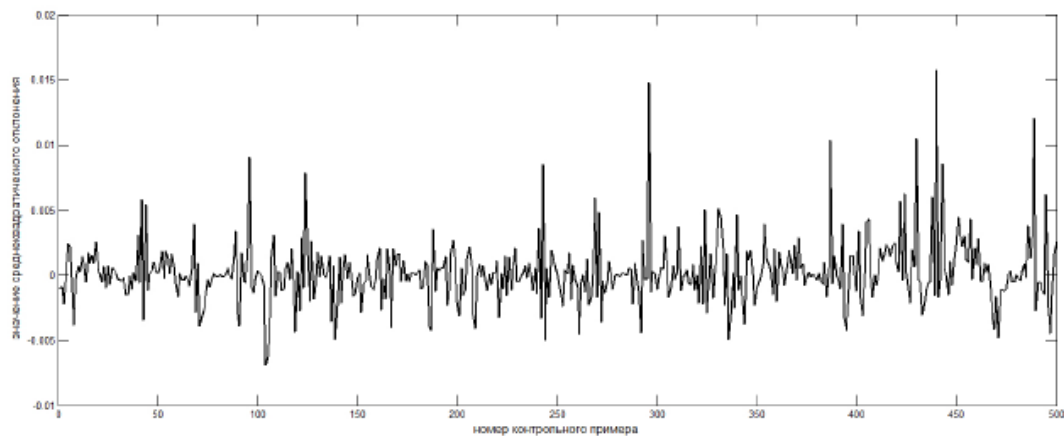


Рис. Ошибка прогнозирования временного ряда, порождённого уравнением Mackey-Glass

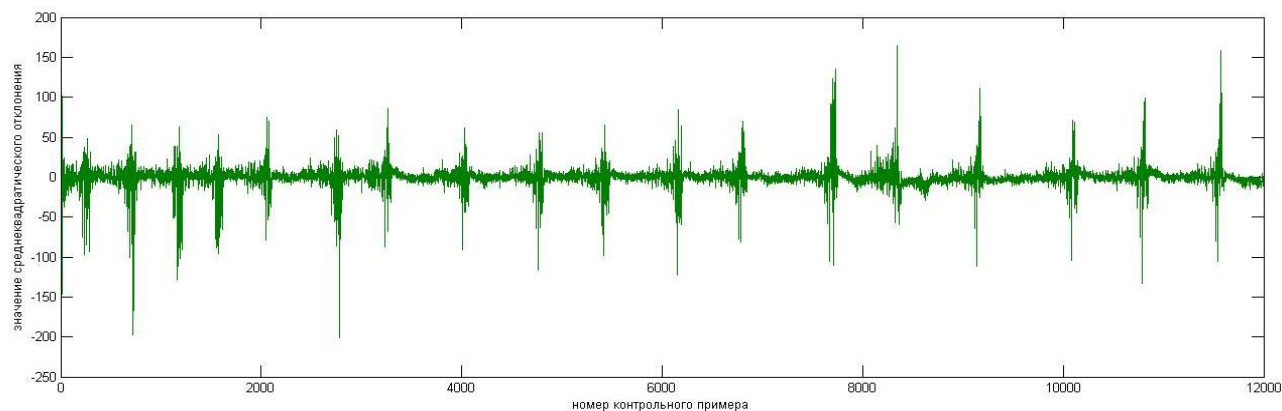
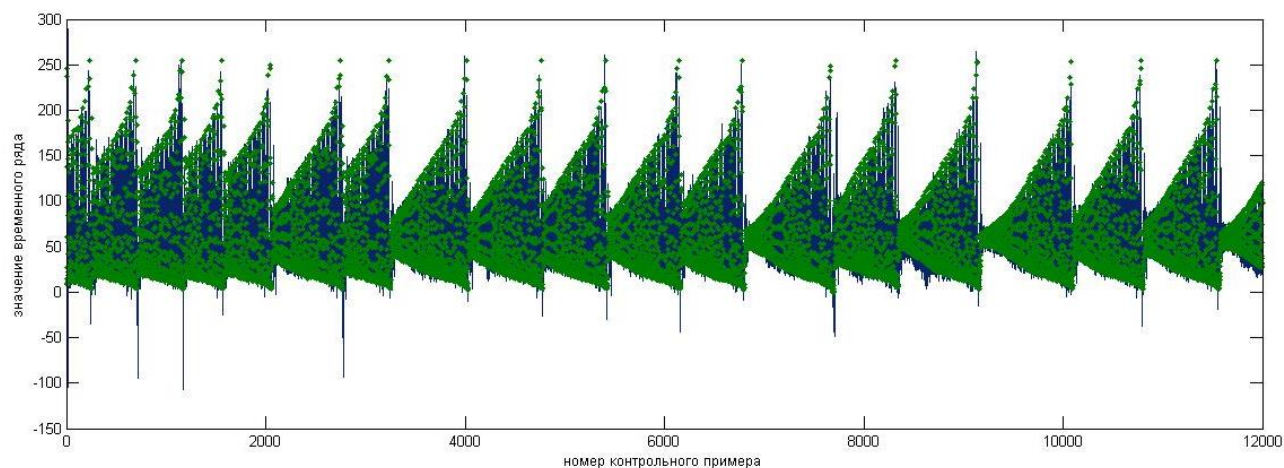
Аппроксимация функции 5 переменных

Сравнение значений ошибки моделирования
при использовании различных методов

Метод	Ошибка
ANFIS	0.007
Mountain method + μ/E оценка (Алгоритм 2)	0.010
Cluster Estimation	0.014
k-средних + μ/E оценка (Алгоритм 1)	0.016
Back-Prop NN	0.02
6-th order polynomial	0.04
Cascaded-Correlation NN	0.06
Auto-regressive Model	0.19
Lineal Predictive Method	0.55

Сравнение с Matlab

Временной ряд представляет собой набор из 25000 наблюдений флуктуаций(колебаний), которые описываются тремя парами нелинейных обыкновенных дифференциальных уравнений и возникают при работе лазера дальнего инфракрасного диапазона.



Сравнение с Matlab

Получили следующие результаты (при 183 кластерах и значении радиуса равном 0.05):

Метод	значение ошибки среднеквадратического отклонения	
	Обучающая выборка	Проверочная выборка
Алгоритм 2	0.0604	0.0830
genfis2	0.1097	0.1426
ANFIS	0.1675	0.1768