# Функции радиального базиса (RBF)

**Функция радиального базиса (RBF)** — функция, значение которой зависит только от нормы аргумента. RBF используются в метрических алгоритмах классификации, в частности в методе потенциальных функций, который (в упрощенном виде) рассматривается в этой статье.

## 1 Постановка задачи

Задана выборка $X^l$, в которой описание каждого объекта является вектором $x_i \in R^n$. Метки классов $y_i$ принадлежат множеству $Y = \{+1, -1\}$. Считается, что каждый объект выборки $x_i$ имеет некоторый «заряд» $\gamma_i$ и создает в пределах окрестности $h_i$ потенциал, вид которого определяется функцией радиального базиса $K(x)$. Таким образом, суммарный потенциал в точке $x \in R^n$ определяется по формуле:

$$\phi(x) = \sum_{y=1}^{l} y_i \gamma_i K(\frac{x_i - x}{h_i}).$$

Знак этого потенциала определяет класс объекта $x$, т.е. классификатор $a(x) = \operatorname{sign} \phi(x)$. Требуется оптимизировать значения параметров $\gamma_i$, $h_i$.

## 2 Ядерные функции

Одной из важных частей алгоритма является выбор функции радиального базиса $K(x)$. В качестве $K$ могут применяться следующие функции (введено обозначение $\|x\| = r$):

1. Гауссиана $G(x) = \exp(-\beta r^2)$, где $\beta > 0$.
2. Ядро логистической регрессии $L(x) = \frac{1}{1+e^{r/\sigma}}$, $\sigma > 0$.
3. Ядро, соответсвующее ньютонову потенциалу $N(x) = \frac{1}{A+r}$, $A > 0$.
4. Модифицированный ньютонов потенциал $N_n(x) = \frac{1}{A+r^n}$, $A > 0$, $n > 0$.
5. Треугольное ядро $T(x) = (1 - r)[r \leq 1]$.
6. Модифицированное треугольное ядро $T_n(x) = (1 - r)^n[r \leq 1]$, $n > 0$.
7. Прямоугольное ядро $P(x) = [r \leq 1]$.
8. Ядро Епанечникова $E(x) = (1 - r^2)[r \leq 1]$.
9. Квартическое ядро $Q(x) = (1 - r^2)^2[r \leq 1]$.
10. Обобщенное квартическое ядро $Q_n(x) = (1 - r^2)^n[r \leq 1]$, $n > 0$.
11. Обратное мультиквадратичное ядро $M(x) = \frac{1}{\sqrt{A+r^2}}$, $A > 0$.
12. Семейство ядер Вендланда $W_{nl}$, представляющее собой многочлены на отрезке $[-1, 1]$ и равные нулю вне его. Они характеризуются двумя параметрами — размерностью пространства $n$ и степенью гладкости $l$. В данной реализации используется первые девять ядер:

| $n$ | $l$ | $W_{nl}$ |
|---|---|---|
| 1 | 0 | $(1-r)_+$ |
| 1 | 2 | $(1-r)_+^3(3r+1)$ |
| 1 | 4 | $(1-r)_+^5(8r^2+5r+1)$ |
| 3 | 0 | $(1-r)_+^2$ |
| 3 | 2 | $(1-r)_+^4(4r+1)$ |
| 3 | 4 | $(1-r)_+^6(35r^2+18r+3)$ |
| 5 | 0 | $(1-r)_+^3$ |
| 5 | 2 | $(1-r)_+^5(5r+1)$ |
| 5 | 4 | $(1-r)_+^7(16r^2+7r+1)$ |

Используется обозначение $F_+ = \max(0, F)$.

# 3 Алгоритм отыскания оптимальных параметров

Считается, что радиусы всех потенциалов равны между собой ($h_i = h \ \forall i = 1, ..., l$). Переменная $h$, а также параметры RBF (если таковые имеются) — структурные параметры метода. Заряды $\gamma_i$ настраиваются согласно алгоритму:

**1.** положить $\gamma_i := 0 \ \forall i = 1, ..., l$.
**2.** повторять
**3.** выбрать случайный элемент из выборки
**4.** если $a(x_i) \neq y_i$ то
**5.** $\gamma_i := \gamma_i + 1$
**6.** пока не выполнен критерий останова

Используются следующие критерии останова алгоритма:

- Ограничение на максимальное количество итераций.
- Доля ошибок на обучающей выборке — если $\sum_{i=1}^{l} [y_i \neq a(x_i)] < \varepsilon$, то алгоритм останавливается. $\varepsilon$ является структурным параметром метода.
- В качестве развития предыдущего пункта — процент ошибок на тестовой выборке $X^k$, которая выделяется из обучающей заранее случайным образом. Доля объектов, которые войдут в тестовую выборку, задается извне.

Поскольку подсчет доли ошибок — весьма трудоемкий процесс, то рационально выполнять его не в каждой итерации алгоритма. Частоту проверок можно изменять извне.

# 4 Вычислительный эксперимент

Для проверки работоспособности алгоритма использовались наборы реальных и синтетических данных.

## 4.1 Реальные данные

В качестве реальных данных использовались данные о сортах ирисов, собранные Фишером. Обучающая выборка состояла из 100 объектов (по 50 объектов каждого из двух классов), обладающих четырьмя вещественными признаками. При использовании всех четырех признаков алгоритм показал очень хороший результат, неправильно классифицировав три объекта. При использовании только первых двух признаков классы глубоко проникают друг в друга; алгоритм работает значительно хуже, неправильно классифицируя около 25 объектов.

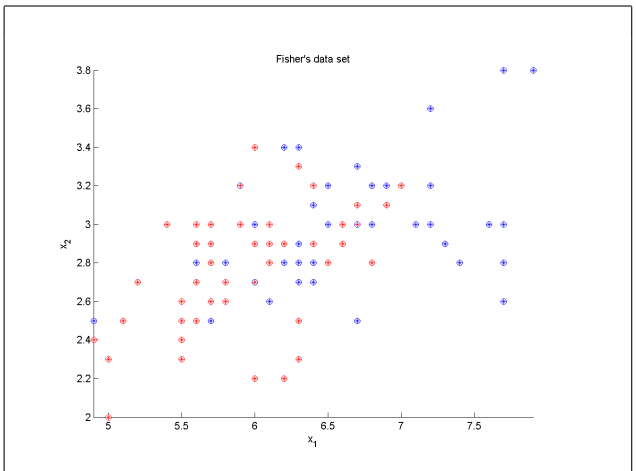


Рис.1. Классификация ирисов, использованы 4 признака.

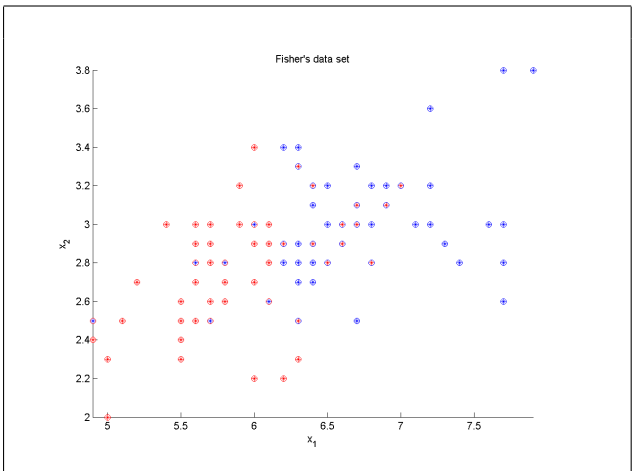


Рис.2. Классификация ирисов, использованы 2 признака.

*По осям графиков отложены значения признаков. Точки — объекты, их цвет точек обозначает принадлежность к тому или иному классу; цвет кружков вокруг точек — результат работы алгоритма.*

Во втором случае возможно нарисовать трехмерный график, изображающий потенциал в различных точках пространства признаков:

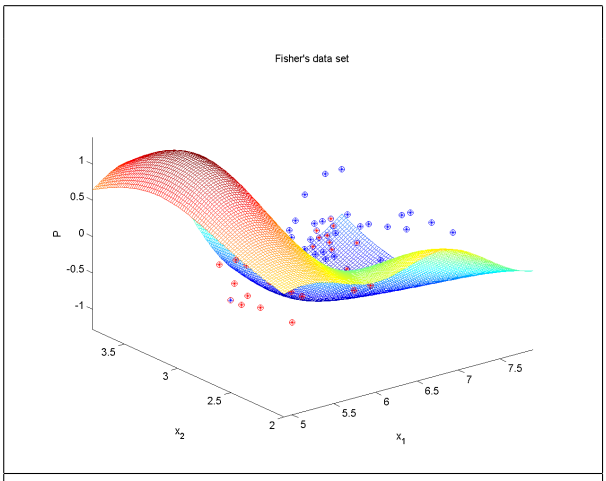


Рис.3. Потенциал, полученный в результате обучения.

## 4.2 Синтетические данные

Синтетические данные представляли собой объекты с двумя признаками из нескольких кластеров. В пределах каждого из кластеров признаки имели нормальное распределение. В зависимости от числа кластеров и параметров распределения задача классификации представляла разную сложность.

### 4.2.1 Легкий случай

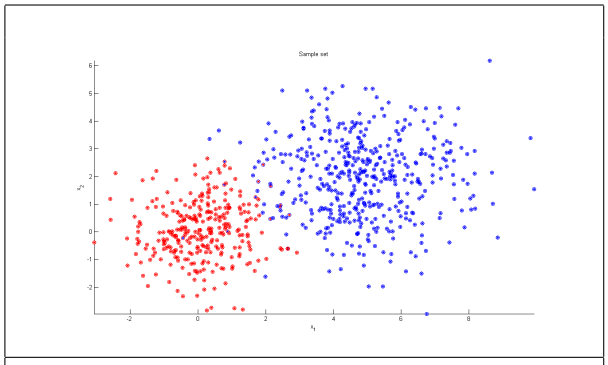

Рис.4. Синтетические данные, легкий случай.

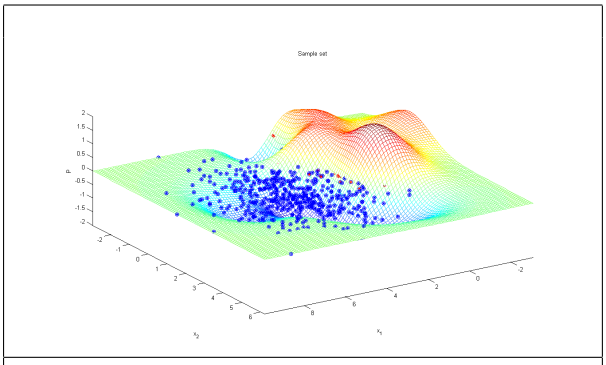

Рис.4а. Потеницал, полученный алгоритмом.

Объекты формируют два кластера со слабым взаимным проникновением. Качество классификации хорошее (меньше 5% неправильно определенных меток классов).

### 4.2.2 Промежуточный случай

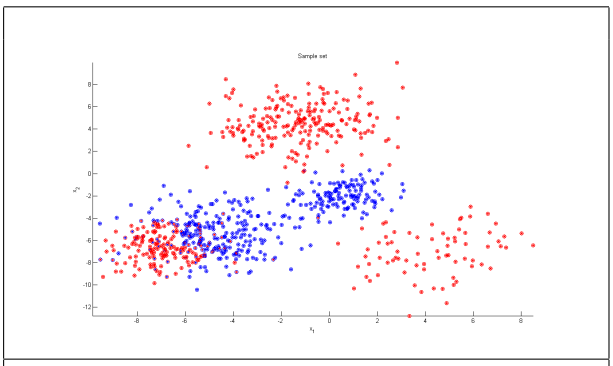

Рис.5. Синтетические данные, промежуточный случай.

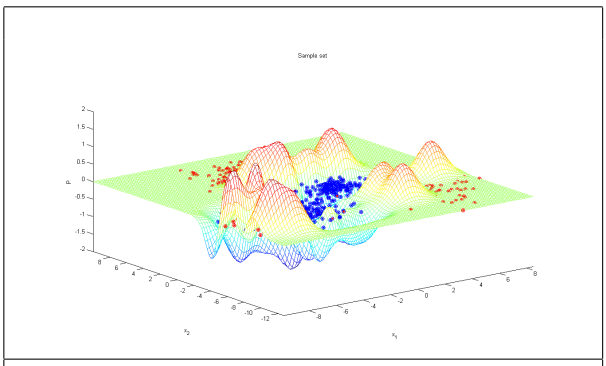

Рис.5а. Потеницал, полученный алгоритмом.

Объекты составляют пять кластеров с существенным проникновением. Неправильно классифицировано около 8% объектов.

### 4.2.3 Сложный случай

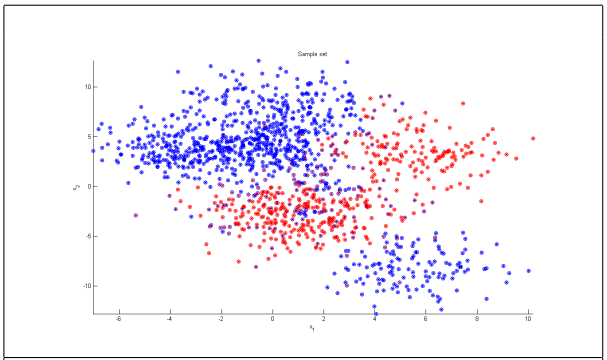

Рис.6. Синтетические данные, сложный случай.

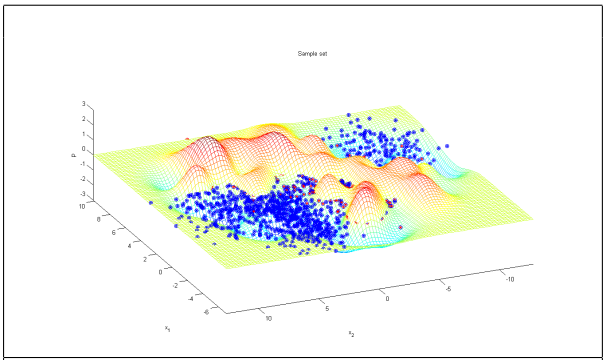

Рис.6а. Потеницал, полученный алгоритмом.

Объекты составляют восемь кластеров с сильным проникновением друг в друга. Неправильно классифицировано около 13% объектов.

# 5 Литература


- К. В. Воронцов. Лекции по метрическим алгоритмам классификации.
- Хардле В. Прикладная непараметрическая регрессия.
- Интерактивная помощь программы MATLAB.
- Bishop C. Pattern Recognition and Machine Learning.