

Понижение размерности

2021

Мария Корлякова

ПЛАН

1. Селекция признаков
2. Редукция системы признаков

Постановка задачи

Задача обучения с учителем

$$X = (x^i, y_i)_{i=1}^l$$

$$a(x) :$$

$$Q(a, X) \rightarrow \min$$

$$a(x) = w_0 + w_1 x^1 + w_2 x^2 + \dots + w_d x^d$$

$$a(x) = \text{sign}(w_0 + w_1 x^1 + w_2 x^2 + \dots + w_d x^d)$$

Этапы формирования модели

Анализ данных и их очистка

Генерация признаков – выявление признаков, которые наиболее полно описывают объект.

Селекция признаков – выявление признаков, которые имеют наилучшие классификационные свойства для конкретной задачи.

Построение модели.

Оценка модели.



процедура представления информации

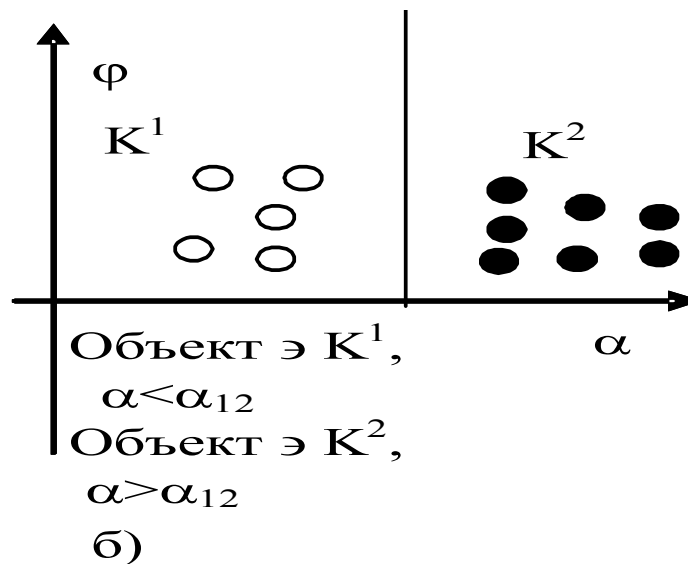
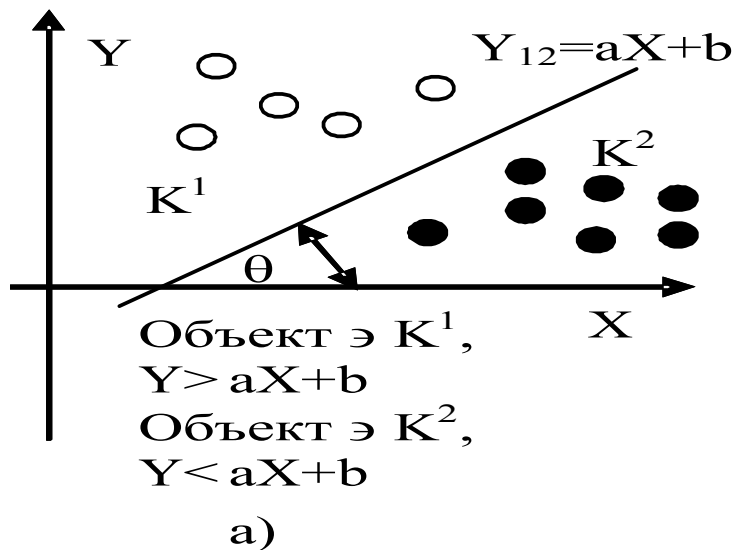
Какие составляющие входной информации следует учитывать?

Какой объем информации необходимо и достаточно сохранять для адекватной работы нейронной сети?

Какие методы следует применять для решения вопросов информативности единиц данных?

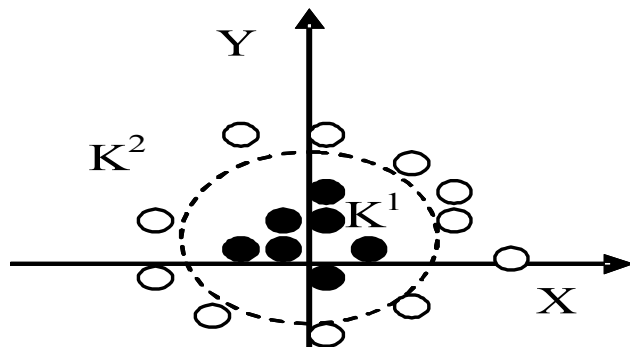


Изменение координат



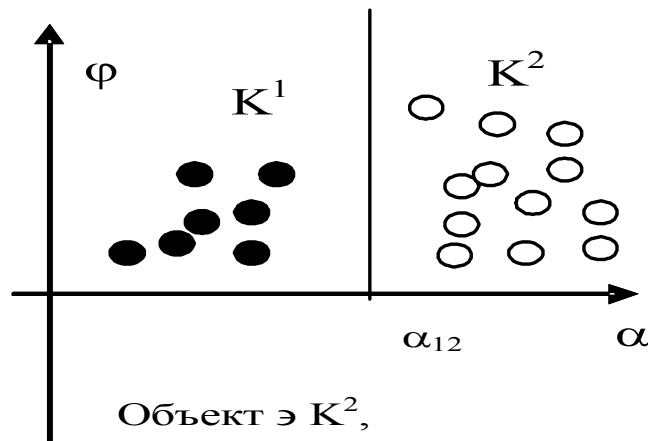
пространство объектов в исходной а) и развернутой на $\pi/2 - \theta$ б)
системе координат.

Изменение координат



Объект $\in K^1$,
 $b^2(Y-b_1)^2 + a^2(X-a_1)^2 < C$
Объект $\in K^2$,
 $b^2(Y-b_1)^2 + a^2(X-a_1)^2 > C$

а)



Объект $\in K^2$,
 $\alpha > \alpha_{12}$
Объект $\in K^1$,
 $\alpha < \alpha_{12}$

б)

пространство объектов в исходной а) и сферической б) системе координат.

Задача селекции признаков

Определение. Процедура выделения из множества признаков меньшего подмножества с наилучшим сохранением информативности для классификации называется селекцией признаков.

Задача селекции признаков

$X \in R^m$ – множество признаков,

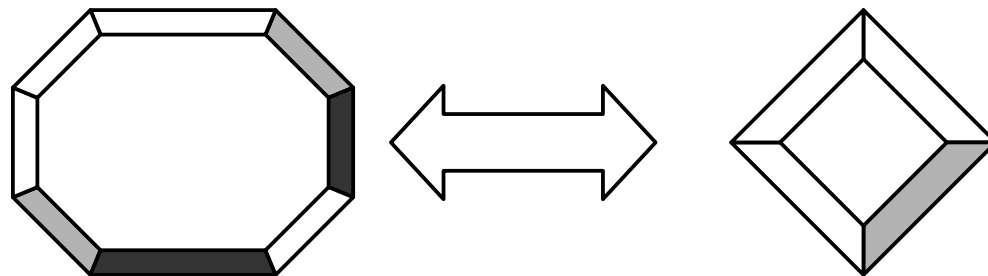
$Y \in R^i$ – множество признаков, которые нужно отобрать в процессе селекции,
причем

$i < m$.

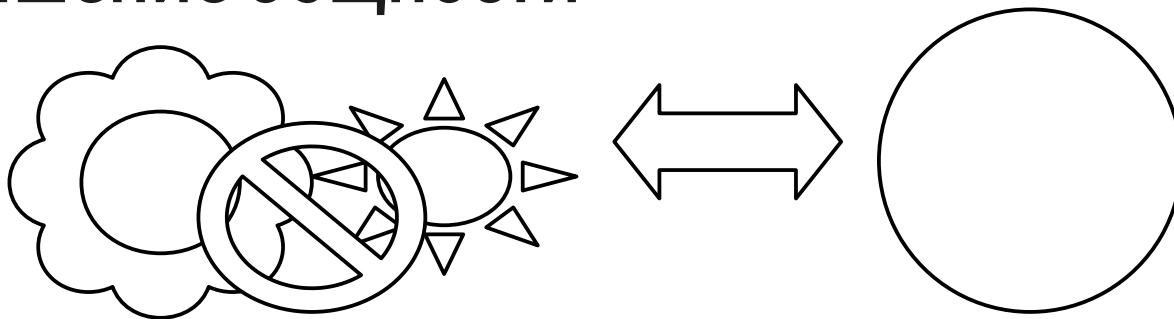
Тогда задача селекции задается следующим образом: $X \rightarrow Y$.

Зачем тратим время на отбор признаков?

Снижение сложности



Повышение общности



Сложность модели

k – число степеней свободы модели


n — порядок порядок вхождения элементов в модель

Общность модели

N – число прецедентов,


k – число степеней свободы модели

$\frac{N}{k}$ характеристика общности модели.



Решение лежит в поиске комбинаций удовлетворяющих гипотезе компактности.

Для отбора из 20 исходных признаков пяти наиболее информативных приходится иметь дело примерно с $15,5 \cdot 10^3$ вариантами.



Типы селекции признаков

Полный перебор

Скалярная селекция

Векторная селекция



Методы селекции

Фильтры

- Статистика - корреляция
- Информативность Шеннона

Оценка через модели

- Оборачивание

Оценка через модели по Регуляризации

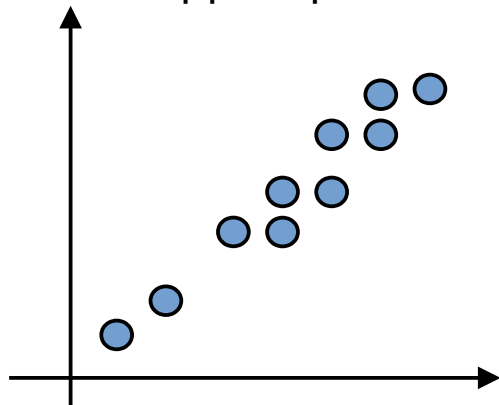


Методы выявления важных признаков — Статистический подход

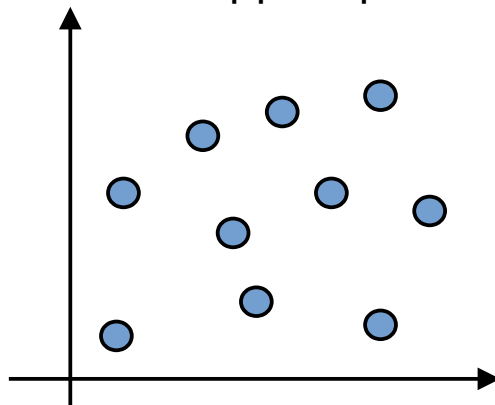
Вычислить корреляцию

$$R_j = \frac{\sum_{i=1}^l (x_{ij} - \bar{x}_j)(y_i - \bar{y})}{\sqrt{\sum_{i=1}^l (x_{ij} - \bar{x}_j)^2 \sum_{i=1}^l (y_i - \bar{y})^2}}$$

Коррелированы



не коррелированы



Информативность : взаимная информация

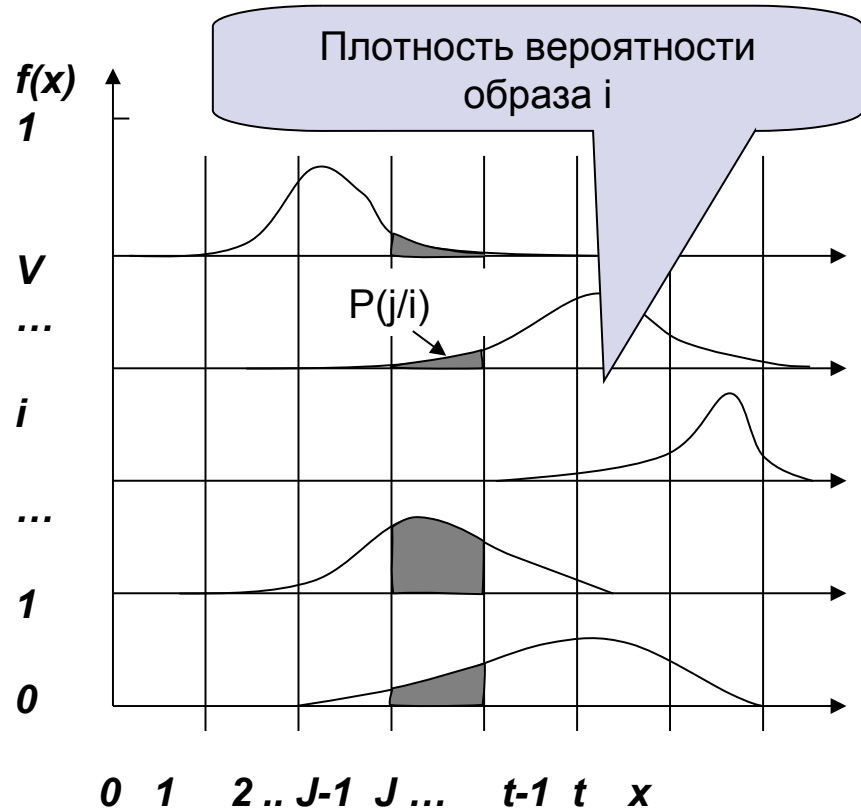
$$MI_j = \sum_{v \in X} \sum_{k \in Y} P(x = v, y = k) \log \frac{P(x = v, y = k)}{P(x = v)P(y = k)}$$

$P(x=v, y = k)$ – доля объектов класса k со значением v для признака x .

$P(y = k)$ – доля объектов класса k .

$P(x=v)$ – доля объектов со значением v признака x .

Информативность



$H(j)=0 \Rightarrow x_j$ – абсолютно разделяющий.

$H(j)<H(m) \Rightarrow x_j$ информативней x_m

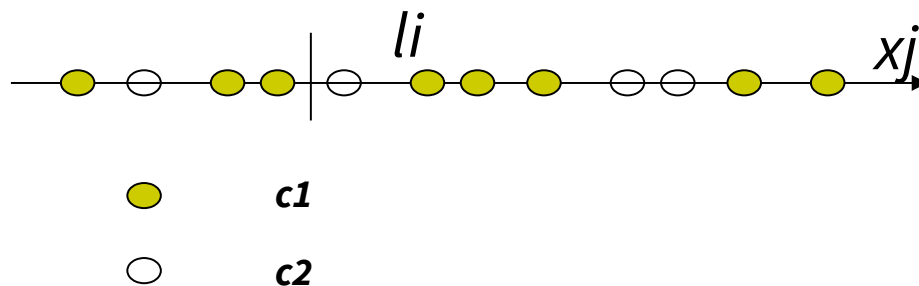
V – число образов

t – число признаков

Дихотомия выборки по признаку

Информативность I

$$I = \min_{l_i} R$$



$$l_i \mid \begin{aligned} R_1 &= m_{1c1} \cdot m_{1c2} \\ R_2 &= m_{2c1} \cdot m_{2c2} \end{aligned}$$

$$R = R_1 + R_2$$

Жадный алгоритм:

- 1) выбрать самый информативный признак :
текущая модель $\underline{X} = \{x \arg \max I(\underline{X}Ux_i)\}$
- 2) пробуем добавить признак x_i из множества оставшихся X^* к \underline{X} и оценим общую информативность $I(\underline{X}Ux_i)$
- 3) $x_j = \arg \max I(\underline{X}Ux_i)$, x_i из X^*
- 4) $\underline{X} = \underline{X} \cup x_j$, $X^* = X^* \setminus x_j$
- 5) $|\underline{X}| < m$, идем к 2)

Модели умеют считать важность признака

Линейная модель

$$a(x) = \sum_{i=1}^n w_i x^i.$$

Деревья

$$Q(X_m, j, t) = H(X_m) - \frac{|X_l|}{|X_m|} H(X_l) - \frac{|X_r|}{|X_m|} H(X_r).$$

Понижение размерности: редукция

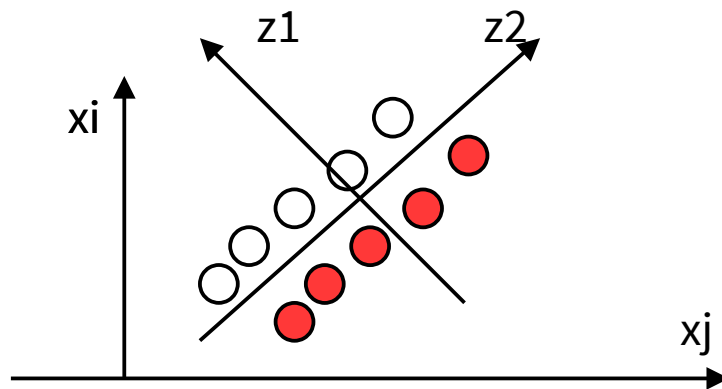
Новые координаты

$$z_{ij} = \sum_{k=1}^n w_{jk} x_{ik}$$

Случайные проекции

$$d > \frac{8 \ln l}{\epsilon^2},$$

l - количество объектов, ϵ - максимальное изменение расстояния между объектами



Метод главных компонент

По n – числу исходных признаков выделить k главных компонент, или обобщенных признаков.

Пространство главных компонент ортогонально.

Модель метода главных компонент основана на допущении, что значения множества взаимосвязанных признаков порождают некоторый общий результат

Метод главных компонент

principal component analysis, PCA

$$Z = XW^T$$

X - матрица "объекты-признаки", где по строкам отложены объекты, а по столбцам - значения признаков,

Z - матрица новых признаков,

W^T - транспонированная матрица весов (W — ортогональна)

$$\|ZW - X\|^2 \rightarrow \min_{Z,W}$$

$$\text{rank}(X) \geq d$$

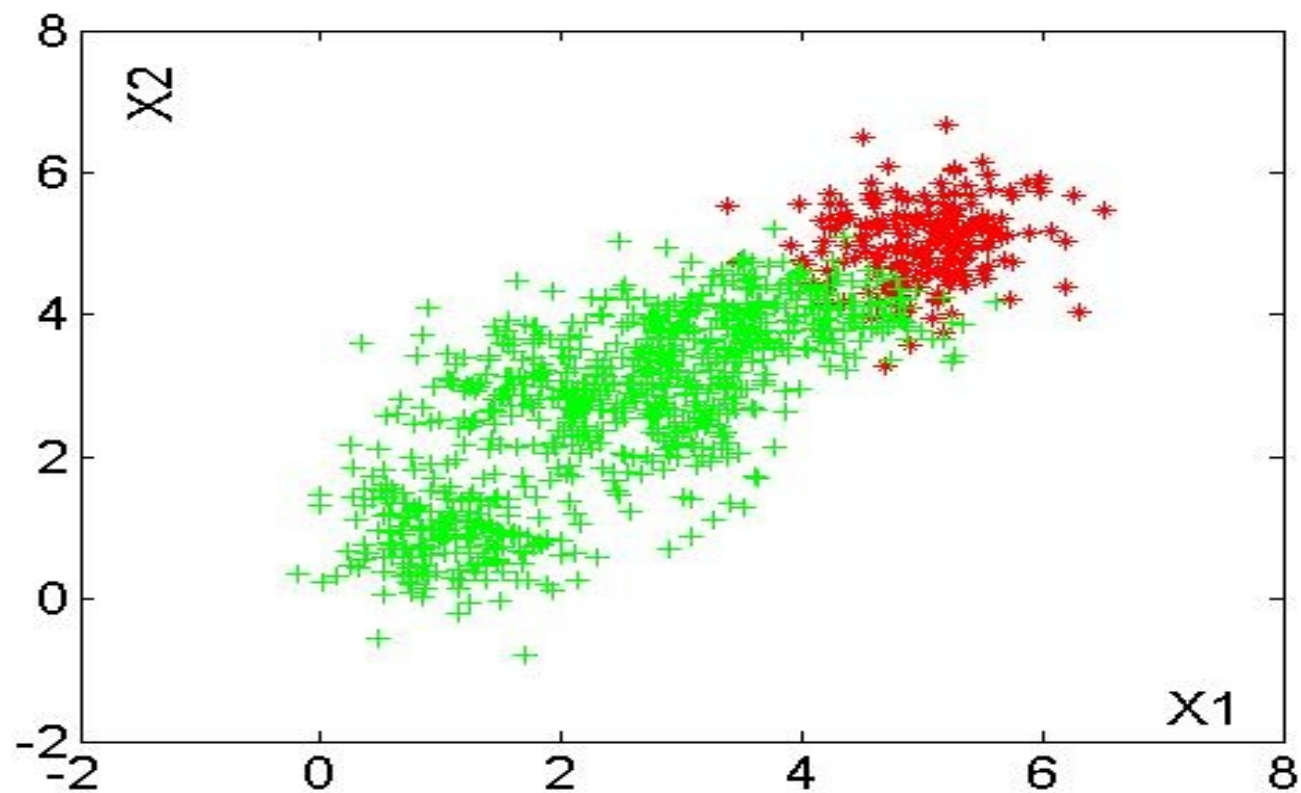
d - число новых признаков,

Метод главных компонент

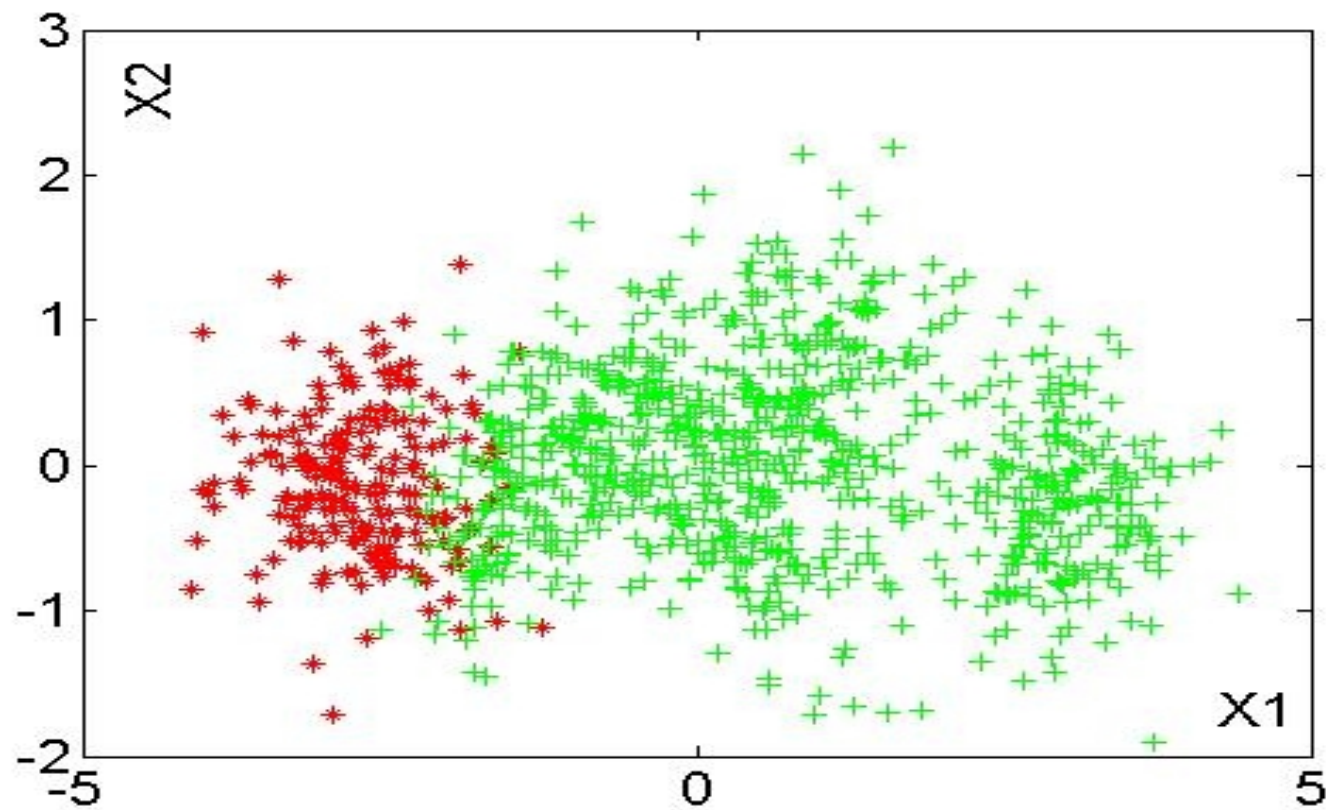
principal component analysis, PCA

- найти собственные значения матрицы $X^T X$;
- отобрать d максимальных;
- составить матрицу W^T ,
- получить $Z = XW$.

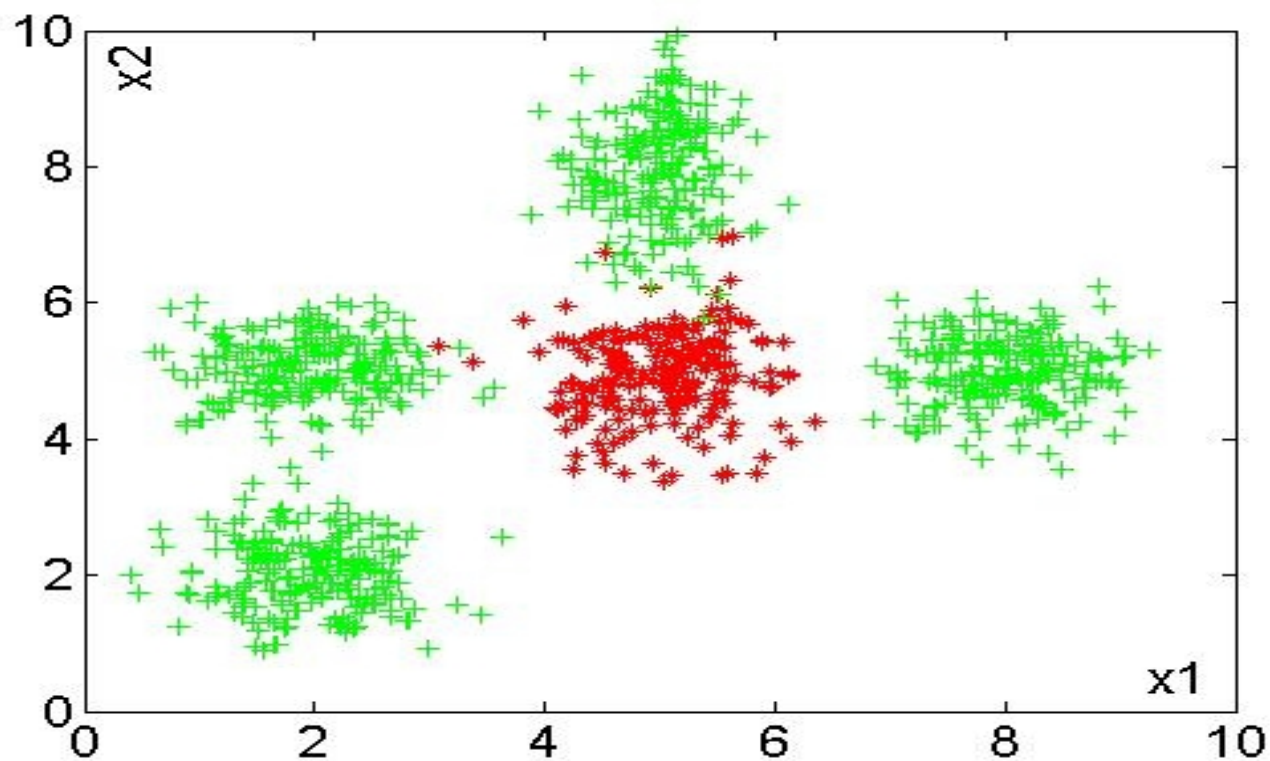
До преобразования



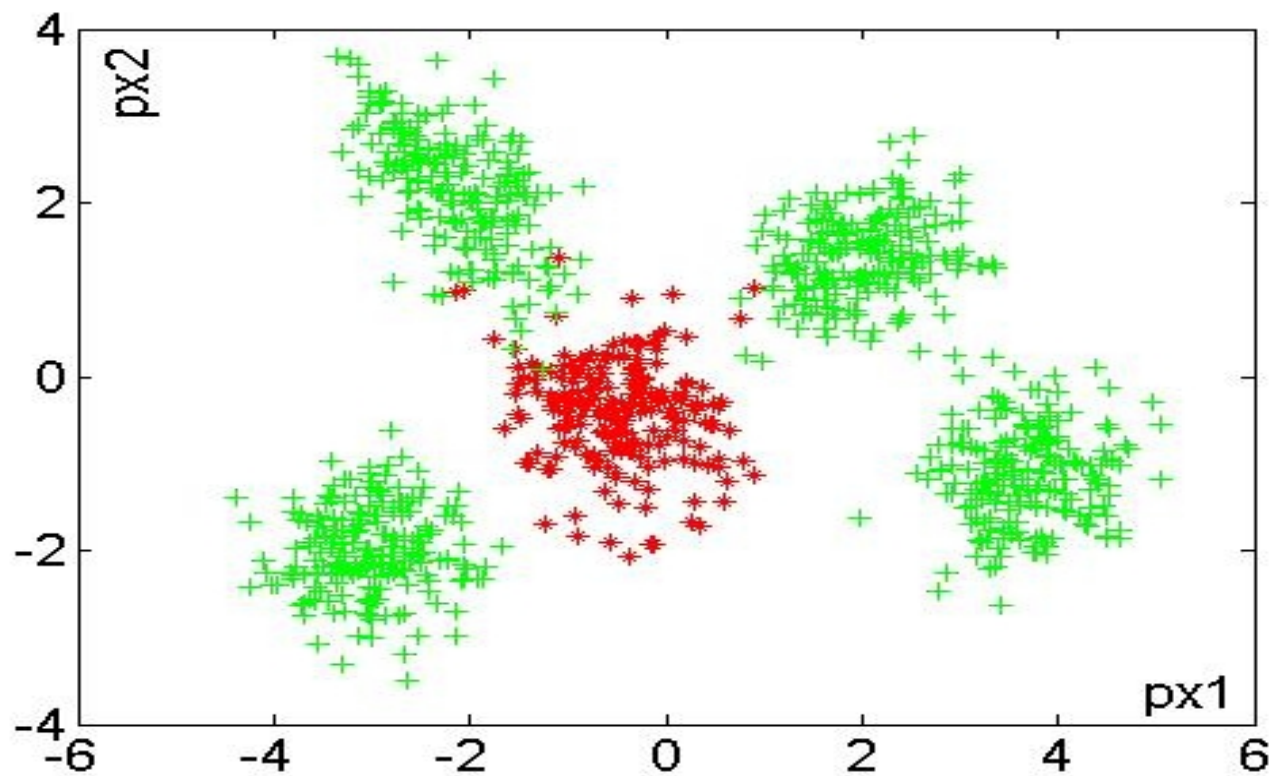
После преобразования



До преобразования



После преобразования



SVD

singular value decomposition, SVD

$$X = UDV^T,$$

U - это собственные векторы матрицы XX^T

столбцы ортогональной матрицы V - собственные векторы матрицы X^TX ,

на главной диагонали диагональной D - собственные значения матриц XX^T и X^TX

При $d = n$ $X = ZW = UDV^T$

SVD

singular value decomposition, SVD

$$X = UDV^T,$$

U - это собственные векторы матрицы XX^T

столбцы ортогональной матрицы V - собственные векторы матрицы X^TX ,

на главной диагонали диагональной D - собственные значения матриц XX^T и X^TX

При $d = n$ $X = ZW = UDV^T$ $W = V^T$ $Z = UD.$

SVD

- найти сингулярное разложение вектора X
- сформировать из столбцов матрицы V , матрицу весов W ;
- получить $Z = XW$.