**Хахин Максим (15 по списку)**

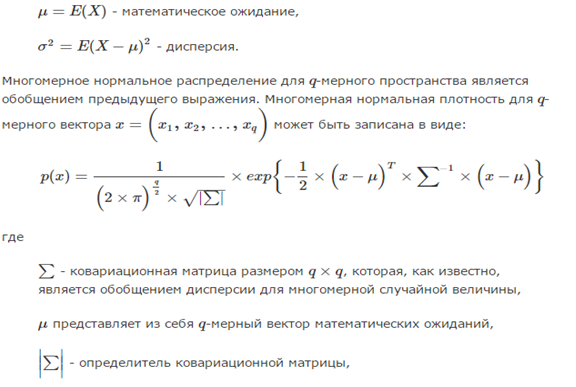
**В среде Weka**

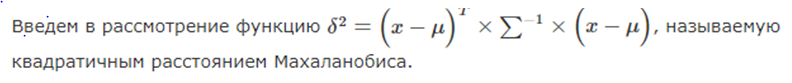
**Задача кластеризации EM-методом**

Дан набор данных ирисы Фишера (см. iris.arff в программе Weka). Столбец class удалить или при запуске EM метода выбрать Ignore attributes и class.

Задание:

1. Выполнить кластеризацию ирисов Фишера EM-методом, указав параметр numClasters= **-1**, а параметр seed=<число, указанное преподавателем – порядковый номер студента в группе, см.список группы> и получить вероятностные характеристики кластеров.
2. Задать произвольные значения ириса <sepallength, sepalwdth, petallength, petalwdth > в виде 4-х вещественных значений, входящих в диапазоны их изменения в обучающей выборке.
3. Используя расстояние Махаланобиса и вероятностные характеристики кластеров, полученные на шаге 1) определить принадлежность к кластеру выбранного на шаге 2) ириса Фишера.





**Задача кластеризации методом k-средних**

Дан набор данных ирисы Фишера (см. iris.arff в программе Weka). Столбец class удалить.

Задание:

1. Выполнить кластеризацию ирисов Фишера методом k-средних, задав количество кластеров, полученное при кластеризации EM-методом, и получить координаты центроидов кластеров
2. Задать произвольные значения ириса <sepallength, sepalwdth, petallength, petalwdth, > в виде 4-х вещественных значений, входящих в диапазоны их изменения в обучающей выборке.
3. Выбрать расстояние и определить принадлежность к кластеру выбранного на шаге 2) ириса Фишера.

**Сравнение**

* Сравнить координаты центроида кластера, полученного методом к-средних, которому принадлежит выбранный Ирис, с координатами математического ожидания кластера, полученному EM-методом, которому принадлежит выбранный Ирис.
* Произвести разбиение EM-методом и методом k-средних на 3 кластера, выбрав Claster mode = Classes to claster evaluation – class. Определить точность построенных моделей.

3.1.

EM-метод

Cluster

Attribute 0 1 2 3 4

(0.28) (0.23) (0.18) (0.15) (0.15)

======================================================

sepallength

mean 6.1614 6.8586 4.7751 5.2824 5.5432

std. dev. 0.4137 0.5228 0.2406 0.2408 0.3159

sepalwidth

mean 2.8547 3.0862 3.1791 3.7039 2.5786

std. dev. 0.2687 0.2891 0.26 0.2857 0.2512

petallength

mean 4.7484 5.7859 1.4194 1.5174 3.863

std. dev. 0.3193 0.4745 0.1691 0.1592 0.3516

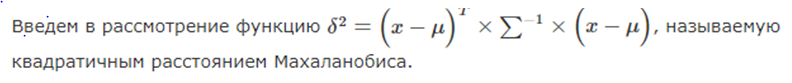
petalwidth

mean 1.5757 2.1327 0.1948 0.3028 1.1696

std. dev. 0.2196 0.2359 0.0556 0.1212 0.1351

**Определить принадлежность кластеру нового объекта:**

**Ирис\*=<sepallength, sepalwdth, petallength, petalwdth >= <5, 3, 4, 3>**

****

**Вычисляем расстояние Махалонобиса от объекта Ирис\* до каждого кластера:**

**d2[0]=(5-**6.1614)2/(0.4137)2+(3-2.8547)2/ (0.2687)2 +(4-4.7484)2/(0.3193)2 + (3-1.5757)2/(0.2196)2 =**55.734**

**d2[1]=(5-**6.8586)2/ (0.5228)2 +(3-3.0862)2/ (0.2891)2 +(4-5.7859 )2/(0.4745)2 + (3-2.1327)2/(0.2359)2 =**40.41**

**d2[2]=(5-**4.7751)2/ (0.2406)2+(3-3.1791)2/ (0.26)2 +(4-1.4194)2/(0.1691)2 + (3-0.1948)2/(0.0556)2 = **2779.76**

**d2[3]=(5-**5.2824)2/ (0.2408)2+(3-3.7039)2/ (0.2857)2 +(4-1.5174)2/(0.1592)2 + (3-0.3028)2/(0.1212)2 =**745.87**

**d2[4]=(5-**5.5432)2/ (0.3159)2 + (3-2.5786)2/ (0.2512)2 +(4-3.863)2/(0.3516)2 + (3-1.1696)2/(0.1351)2 =189.48

**<sepallength, sepalwdth, petallength, petalwdth >= <5, 3, 4, 3> - ирис принадлежит 1-му кластеру**

**Final cluster centroids:**

**Cluster#**

**Attribute Full Data 0 1 2 3 4**

**(150.0) (27.0) (26.0) (27.0) (50.0) (20.0)**

**===============================================================================**

**sepallength 5.8433 6.0296 5.55 6.9667 5.006 6.55**

**sepalwidth 3.054 2.7556 2.5808 3.137 3.418 3.05**

**petallength 3.7587 4.9444 3.9269 5.8852 1.464 4.805**

**petalwidth 1.1987 1.7037 1.2 2.2 0.244 1.55**

**Вычисляем евклидово расстояние или расстояние городских кварталов**

****

**от объекта Ирис\* до цендроида каждого кластера:**

**d2[0]=** 3.5147

**d2[1]=** 2.8423

**d2[2]=** 4.7656

**d2[3]=** 5.7160

**d2[4]=** 3.8549

**<sepallength, sepalwdth, petallength, petalwdth >= <5, 3, 4, 3> - ирис принадлежит 1-му кластеру**