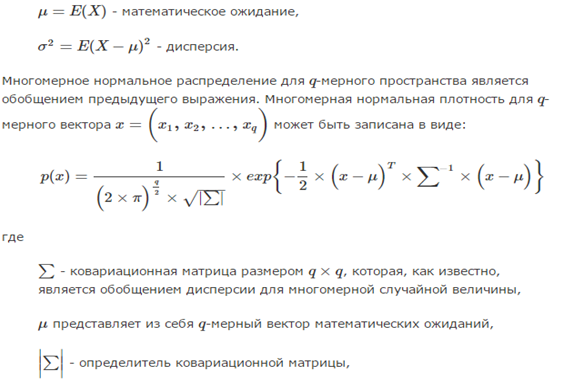
**В среде Weka**

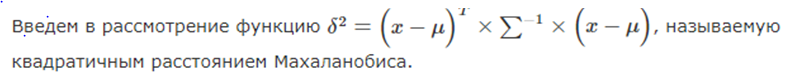
**Задача кластеризации EM-методом**

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 = 100 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**

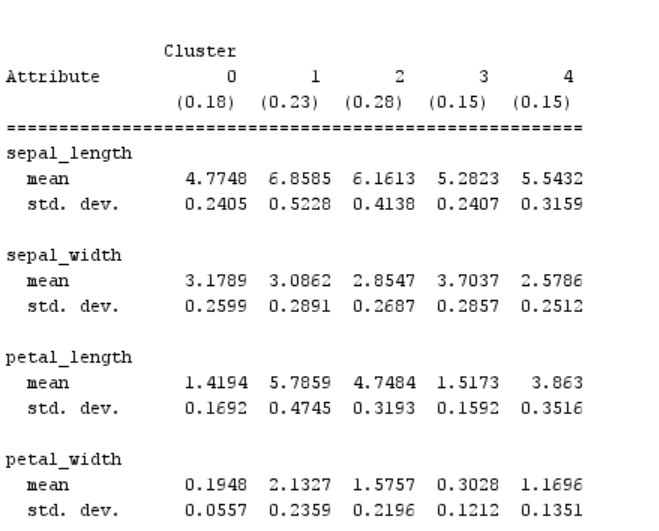
Дан набор данных ирисы Фишера (см. iris.arff в программе Weka). Столбец class удалить.

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.

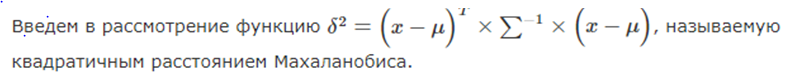
3.1.

EM-метод



**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 **–** 4.778)2/ (0.2405)2+(3 – 3.1789)2/ (0.2599)2 +(4 – 1.4194)2/ (0.1692)2 + (3 – 0.1948)2/ (0.0557)2 = **2770.3397**

**d2[1]** = (5 **–** 6.8585)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 = **40.4091**

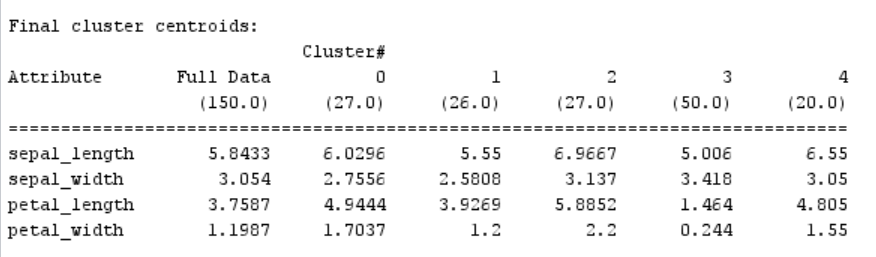
**d2[2]** = (5 **–** 6.1613)2/ (0.4138)2+(3 – 2.8547)2/ (0.2687)2 +(4 – 4.7484)2/ (0.3193)2 + (3 – 1.5757)2/ (0.2196)2 = **55.7289**

**d2[3] =** (5 **–** 5.2823)2/ (0.2407)2+(3 – 3.7037)2/ (0.2857)2 +(4 – 1.5173)2/ (0.1592)2 + (3 – 0.3028)2/ (0.1212)2 = **745.8877**

**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 = **189.4842**

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

3.2  
Метод к-средних



**Вычисляем евклидово расстояние или расстояние городских кварталов**

****

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

**d2[0]=** 1,9215

**d2[1]=** 1,9297

**d2[2]=** 2,8426

**d2[3]=** 3,7685

**d2[4]=** 2,2705

**<sepallength, sepalwdth, petallength, petalwdth >= <5, 3, 4, 3> - ирис принадлежит 0-му кластеру**