***(Аномальные объекты-выбросы).***

Объекты выборки описывается набором из разнотипных (номинальных и количественных) признаков . Множество целочисленных значений градаций номинального признака заменяется на значение частоты встречаемости

где — число объектов с градацией . Для признаков используется нормирование значений в [-1;1].

Требуется найти центр гипершара и радиус *r*, охватывающего всю выборку кроме аномальных объектов-выбросов. Минимизация радиуса гипершара *r* и суммы штрафов за выход из шара определяется по критерию

где расстояние по метрике Чебышева, функция потерь

,

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

***Anomalous Objects (Outliers)***

A dataset is described by a set of heterogeneous (nominal and quantitative) features . The set of integer values of the gradations of a nominal feature is given by . Each nominal feature is replaced by its frequency of occurrence

where is the number of objects with the gradation . For all features, values are **normalized to the range** [−1,1][-1,1].

The goal is to find the **center** and **radius** of a hypersphere that encloses the entire dataset **except for anomalous objects (outliers).** The **minimization of the hypersphere's radius** and penalty for objects outside the sphere is defined by the criterion:

where distance by Chebyshev metric, loss function

,

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*Uzbek\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**Evklid va Chebyshev metrikalari hamda 7 ta yo’qotish funksiyasi bo’yicha hisoblaymiz. Shunda 14 ta qiymat chiqarilib solishtiriladi.**

Metrikalar: Evklid, Chebyshev

Yo’qotish funksiyalari:

S = np.array([

        [80, 100, 32, 12, **30**],

        [76, 200, 28, 14, **30**],

        [84, 200, 30, 10, **60**],

        [ 5, 100, 99, 90, **60**],

        [82, 100, 30,  8, **60**],

        [ **1,   0,  1,  1,  0**]

])

1. Berilgan to’plamni tayyorlab olamiz. Ya’ni har bir obyektning **miqdoriy** alomatlarini

[-1, 1] oraliqqa tushirib olamiz. Normallashtirish:

80=> 2 \* (80-5)/(84-5) – 1; 32=> 2 \* (32-28)/(99-28)-1; 12=> 2 \* (12-8)/(90-8)-1

1. **Nominal** alomatlarini esa quyidagi mezon bo’yicha son ko’rinishiga o’tkazamiz:

100=> 3/(3+2); 200=> 2/(3+2)

3.1 taqsimot zichligi uchun (radius) ni hisoblab olamiz. Bunda KNN algoritmidan foydalanamiz. K=3,4,5 … n holatlar uchun larni ko’rib chiqamiz. Misol uchun har bir obyektning k=3 ta eng yaqin qo’shnisini topib, shu holatdagi radiusini topamiz. Topilgan radiuslarni umumiy soniga bo’lish orqali ni hisoblaymiz.

3.2 topilgan asosiyda

Taqsimot zichligidan foydalanish uchun gipersharni radiusi ni hisoblash kerak bo’ladi. Buni KNN algoritmidan foydalangan holatda toppish mumkin. Avvaliga