TARTU RIIKLIK ÜLIKOOL

Arvutusmatemaatika kateeder

R Ü H M I T A M I S M E E T O D I T E S T.

Diplomitöö.

Töö teostaja: V e s k i, Anu.
Matemaatikateaduskonna
V kursuse üliõpilane.

Töö juhendaja: dots. L. V õ h a n d u.

Tartu 1969.

I ÜLEVAADE RÜHMITAMISMEETODITEST.

Paljutunnuselise objektidesüsteemi rühmadeks jaotamist on aastaid uuritud. Kirjanduses on esitatud arvukalt rühmitamismeetodeid, kuid siiani esitatud meetodeist pole ükski täielikku tunnustust leidnud, vt./3/.

Rühmitamismeetodite sagedase kasutamise põhjuseks on suhteliselt lihtne matemaatiline aparatuur, mistõttu võivad neid edukalt kasutada ka need inimesed, kellel puuduvad erilised eelteadmised matemaatikast ja statistikast. Rühmitamine annab võimaluse ükskõik millises uurimisvaldkonnas uuritavate objektide kohta saadud andmeid viia edasise analüüsi jaoks ülevaatlikumale kujule. Selleks jaotatakse uuritav objektikogum rühmadeks nii, et iga rühma liikmed oleksid omavahel võimalikult sarnased, vt. /5/.

Et olemasolevaid meetodeid edukalt kasutada, on vajalik nende võrdlemine ja liigitamine. Kirjanduses leiduvaid rühmitamismeetodeid võib jagada kahte suurde rühma:

1) induktiivsed meetodid,

2) deduktiivsed e. hierarhilised meetodid.

Induktiivsete meetodite puhul algab analüüs üksikobjektide võrdlemisest ning rühmadeks ühendamisest. Deduktiivsete meetodite juures võetakse vaatluse alla terve objektikogum ja jaotatakse teda seni, kuni saadakse homogeensed osad, mis defineeritakse kui rühmad. Deduktiivsete meetodite tuletamisel on jõutud tunduvalt kõrgema viimistletuseni ja arvutuste keerukuseni, vt. /1/. Puuduseks on aga neile küllalt suur arvutustööde maht, sest analüüs nõuab kõigi objektipaaride vaheliste seoste välja arvutamist. Teiselt poolt on hierarhilised meetodid sobivad rakendada vaid selgesti avalduva heterogeensuse korral. Kuna praktikas on hierarhilist rühmitamist võimalik suhteliselt vähem rakendada, on viimasel ajal pööratud suuremat tähelepanu induktiivsete meetodite arendamisele.

Täielikud induktiivsed rühmitamismeetodid võivad koosneda põhimõtteliselt neljast osast:

1) algrühmade määramine,
2) uute objektide lisamine olemasolevatele rühmadele või olemasolevate rühmade omavaheline liitmine,
3) rühmitamise lõpetamine (seejuures võib osa objekte jääda rühmitamata ja moodustada üheobjektilisi rühmi),
4) ümberrühmitamine pärast rühmade moodustamist (võimalike valerühmituste kõrvaldamiseks).

Kõik induktiivsed rühmitamismeetodid sisaldavad esimest ja teist osa. Mõnede meetodite puhul puudub kas kolmas või neljas osa või puuduvad mõlemad osad, vt. /2/.

Rühmitamisele asumisel on oluline sobiva sarnasuse mõõdu valik (eriti hierarhiliste meetodite puhul).

Üheks sagedamini kasutatavaks mõõduks on eukleidiline kaugus:

$$d_{\alpha\beta} = \sum_i (x_{i\alpha} - x_{i\beta})^2.$$

Kaalutud Eukleidilist kaugust

$$d_{\alpha\beta} = \sum_i k_i (x_{i\alpha} - x_{i\beta})^2.$$

/6/

on oma töödes kasutanud Bonner (1962), Sebestyen (1962) /10/, Ball ja Hall (1964) jt. Lisaks eukleidilisele kaugusele on kasutusel veel terve rida mitte- eukleidilisi kaugusi.

Teiseks kasutatavamaks sarnasuse mõõduks objektide võrdlemisel on korrelatsioonikoefitsient:

$$R_{ij} = \frac{E[(X_i - EX_i)(X_j - EX_j)]}{\sqrt{DX_i \cdot DX_j}}, \text{ vt. /4/.}$$

Nii eukleidilist kaugust kui ka korrelatsioonikoefitsienti saab kasutada ka rühma raadiuse ligikaudse mõõduna.

Lisaks neile on kasutusel veel terve rida sarnasuse näitajaid, mis põhinevad informatsiooni mõistel, vt. /5/.

Olemasolevaid induktiivseid rühmitamismeetodeid võib jagada kolme suurede rühma:

1) meetodid, mis nõuavad kõigi objektipaaride vahelise korrelatsiooni või kauguse leidmist,
2) meetodid, mis võtavad üksikobjektid vaatluse alla järjekorras,

3) meetodid, mis opereerivad objektikogumitega.

Enamus kirjanduses esitatud meetodeist kuuluvad esimesse rühma. Vaatleme mõningaid meetodeid igast rühmast.

1962.aastal esitas Needham meetodi /11/, mis algab kõigi objektide-vahelise kauguste maatriksi leidmisega. Seejärel arvutatakse keskmine kaugus - nn. lävi. Kauguste maatriksi elemendid muudetakse nullideks või ühtedeks sõltuvalt sellest, kas lävi ületatakse või mitte. Selliselt teisendatud maatriksi i-s veerg annab i-ndale objektile lähedaste objektide loetelu. Kui võrreldakse i-ndat ja j-ndat veergu, siis uue rühma moodustavad i-ndale ja j-ndale objektile ühised lähedased objektid. Protseduuri korratakse seni, kuni leitakse kõik rühmad vastavalt ette antud arvule.

Sokal ja Michner ühendavad omavahel objektid, millel on suurim paariviisiline korrelatsioon. Järgmisena liidetakse objekt, millel on suurim keskmine korrelatsioon antud rühma liikmetega. Rühma lõpetamise kriteerium saadakse korrelatsiooni vähenemisest. "Oluline" langus määratakse empiiriliselt, vt. /12/.

Induktiivsete rühmitamismeetodite üldiseks puuduseks ongi rühma lõpetamise kriteeriumite puudumine. Täpsed kriteeriumid on antavad vaid selgesti avalduva heterogeensuse korral.

Terve sarnasusmaatriksiga opereerivaid meetodeid on esitanud veel Bonner (1964), Goodall (1966), Sebestyen (1965) jt.

Tunduvalt väiksema töömahuga on meetodid, mis vaatlevad objekte järjekorras. Lihtsaima seda tüüpi meetodi andis Bonner (1964) ning täiuslikuima Hyvarinen (1960), vt. /2/.

Bonner'i meetod /2/.

Juhuslikult valitakse välja üks objekt. Edasi leitakse ülejäänud objektide kaugused valitud lähteobjektist. Küllalt lähedased objektid (mis on lähemal empiiriliselt ette antud lävest), liidetakse lähteobjektile ja moodustatakse uus rühm. Uus lähteobjekt järgmise sammu jaoks valitakse jällegi juhuslikult ülejäänud objektide seast. Küllalt lähedased objektid rühmitatakse analoogiliselt esimese sammuga. Rühmitamist teostatakse seni, kuni kogu objektimassiiv on ammendatud. Seejärel hinnatakse saadud rühmi. Rühmade sobivuse mõõduks on nende eraldatus.

Hyvarinen'i meetod /13/.

Esmakordselt toodi sisse "tüüpilisuse" mõõt. Lähteobjektiks valitakse "tüüpiline" objekt, mis annab suhteliselt vähe informatsiooni. Sellel eeldusel asub lähteobjekt suure tõenäosusega sellises piirkonnas, mis on tihedalt asustatud. Ülejäänud objektid, mis asuvad empiiriliselt ette antud lävest lähemal, liidetakse antud objektile ja moodustatakse rühm. Leitud rühma objektid kõrvaldatakse edasisest vaatlusest. Ülejäänud objektide seast valitakse jällegi uus "tüüpiline" lähteobjekt, moodustatakse uus rühm jne. Peale algset valikut rühma tsentreid (raskuskeskmeid) enam ümber ei muudeta. Suurte objektimassiivide korral on see meetod küllalt töömahukas. Saadud rühmade sobivus sõltub empiiriliselt ette antud sarnasuse piirist.

Seoses rühmitamismeetodite laialdase kasutamisega, on tekkinud vajadus suurte objektimassiivide rühmitamise järele. Suurt töömahtu on püütud vähendada sellega, et ei leita kõigi objektide vahelisi kaugusi. Paljude praktiliste ülesannete korral on leitud, et terve sarnasusmaatriksi defineerimine üle kogu obektimassiivi pole vajalik, vaid piisab kui defineerida sarnasusmaatiks teatud homogeensetele objektirühmadele, vt. /6/.

Selliseid meetodeid on esitatud terve rida ja nad moodustavad kolmanda suure rühma induktiivsetest rühmitamismeetoditest.

Nende meetodite puhul eeldatakse á priori ette antavat rühmade arvu. Rühmade arv võib rühmitamise käigus muutuda või mitte.

Meetodite kirjeldamisel kasutatakse ruumilisi mudeleid. Olgu antud n objekti, mida iseloomustavad m tunnust. Siis objektid kujutavad endast teatud punktikogumit m- dimensionaalses ruumis. Objektide omavaheliseks võrdlemiseks kasutatakse Eukleidilist kaugust, rühmi iseloomustatakse nende raskuskeskmega. Vanim selline töö on Forgy'lt (1965).

Objektikogum jaotatakse k rühmaks selliselt, et valitakse k objekti rühmade tsentreiks ja kõik objektid liidetakse rühmadele, milliste raskuskeskmeile on nad lähemal. Analoogilise meetodi kirjeldas ka Jancy (1966). Paindlikuma kuju andis meetodile MacQueen 1966.aastal. Esialgseiks rühma tsentreiks valitakse juhuslikult k objekti. Iga objekt liidetakse olemasolevaist rühmadest sellega, mille tsenter (raskuskese) on lähim. Objekte vaadeldakse järjekorras. Pärast objekti rühmale liitmist leitakse rühma uus raskuskese. Rühmade raskuskeskmed muudavad oma asukohti semi, kuni objektide rühmitamine on lõppenud. Kui kahe rühma raskuskeskmed tulevad teine-teisele lähemale, kui antud ülesande puhul on ette lubatud, liidetakse nad ning rühmade arvu k vähendatakse ühe võrra. Kui mõni objekt asub olemasolevaist rühmadest küllalt kaugel, moodustab ta uue, üheobjektilise rühma ja k suureneb ühe võrra.

MacQueen kasutas saadud tulemuste parandamiseks peale rühmitamise lõpetamist ka ümberrühmitamist. Selleks võrreldi kõiki objekte rühmitamise lõpus uuesti leitud rühmade raskuskeskmetega. Võrdlemise eesmärgiks oli rühmitamise käigus saadud väikeste rühmade ümberjaotamine.

II Võhandu meetod.

Induktiivsete rühmitamismeetodite kasutamisega saab arvutustööde mahtu tunduvalt vähendada. Kuid vaatamata sellele on koostatud rühmitamisprogrammide töö rohkearvuliste võrdluste tõttu küllalt aeglane. Praktikas esinevate suurte objektimassiivide rühmitamisel muutub olukord veelgi keerukamaks. Ülesannete lahendamist hakkab piirama masina mälu maht ning arvutusteks kuluv pikk aeg.

Nagu TPI Arvutuskeskuse praktika on näidanud, on suhteliselt väikese arvutustööde mahuga nn. puu-tüüpi rühmitamismeetodid. Selleks, et n objektile joonistada minimaalsete ühendusteedega puud, on tarvis välja arvutada kõigi objektide vahelised kaugused. Neid kaugusi on $\frac{n(n-1)}{2}$. Minimaalsete ühendusteedega puu välja joonistamisel aga ei kasutata kaugeltki kõiki leitud kaugusi. See viib mõttele, et ilmselt on võimalik objektimassiivi struktuuri eelneva analüüsiga puu-tüüpi rühmitamiseks vajalike kauguste arvu tunduvalt vähendada.

Ühe võimaluse annab L.Võhandu meetod. Siin kasutatakse ära asjaolu, et põhilise informatsiooni iga objekti kohta annab antud objekti lähim naaber, vt. /3/.

Järelikult võime iga objekti rühmitamisel piirduda tema paari lähima naaber-objekti vaatlemisega, kaugemad objektid võime vaatluse alt välja jätta. Et meie eesmärgiks on arvutuste mahu vähendamine, peaksid vaatluse alla tulevad objektid olema kogu objektimassiivi ulatuses üksteisest küllaltki erinevad. Seega tuleks leida objektimassiivist üles omavahel küllalt erinevad objektid – nimetame neid koordinaatobjektideks ning seejärel leida igale objektile u ... n lähimat koordinaatobjekti, mis määravad kindlaks objekti asukoha m-dimensionaalses tunnuste ruumis.

Iga u sellist koordinaatobjekti määrab selles m-dimensionaalses ruumis mingi hüperkolmnurga. Selliste hüperkolmnurkade arv sõltub objektimassiivis leiduvate koordinaatobjektide arvust. Kui me igale objektile leiame u lähimat koordinaatobjekti, määrame me sellega kindlaks antud objektile lähima hüperkolmnurga, Omavahel lähedased objektid koonduvad seejuures ilmselt ühise hüperkolmnurga sisse või vahetusse lähedusse.

Nüüd võime objektimassiivist üles otsida need objektid, millel on ühised lähimad koordinaatobjektid ja leida nende objektide omavahelised kaugused. Sellega oleme $(n \times n)$ kauguste maatriksist välja arvutanud vaid teatud " tükid" omavahel lähedaste objektide vaheliste kaugustega. Leitud kauguste abil saab täieliku minimaalsete ühendusteedega puu välja joonistada ning seejärel leida rühmad. Töömaht on tunduvalt vähenenud.

Käesolevas diplomitöös piirdutakse kõigi kooordinaatobjektide ning igale objektile u lähima koordinaatobjekti leidmisega. Objektide edasine rühmitamine ühiste koordinaatobjektide järgi jäetakse vaatluse alt välja.

Koordinaatobjektide leidmisel kerkib üles kaks küsimust. Milline on koordinaatobjektide arv konkreetse objektimassiivi korral? Milline peab olema objektide vaheline kaugus, et neid lugeda koordinaatobjektideks?

Koordinaatobjektide arvu h_{max} ja koordinaatobjektide vahelise minimaalseima kauguse d_{min} vahel on ilmselt seos olemas, sest mida vähem me lubame koordinaatobjekte, seda kaugemal nad peavad üksteisest asuma. Täpset funktsionaalset seost nende vahel on raske määrata, sest see sõltub objektimassiivi struktuurist. Sellise heuristilise meetodi rakendamisel on konkreetse objektimassiivi korral otstarbekas piirduda suuruste h_{max} ja d_{min} empiirilise ette andmisega.

Koordinaatobjektide arvu h_{max} ette andmisel tuleb arvestada, et suurus h_{max} on nii alt kui ülalt tõkestatud. Alumiseks tõkkeks on säiluv informatsioon. Mida väiksem on koordinaatobjektide arv, seda harvemini nad paiknevad objektimassiivis, seda vähem informatsiooni annavad objekti lähimad koordinaatobjektid, Selle piiri kindlaks tegemiseks tuleb konkreetset ülesannet lahendada mitme intuitiivselt ette antava koordinaatobjektide arvuga. Ülalt piirab koordinaatobjektide arvu objektide arv n. Et meie eesmärgiks on rühmitamiseks vajalike arvutuste vähendamine, huvitab meid võimalikult väike koordinaatobjektide arv

$$h_{max} \ll n.$$

Arvutatavate kauguste arv koordinaatobjektide otsimisel sõltub objektimassiivi struktuurist ning esimese koordinaatobjekti valikust. Juhul, kui enamus otsitavaid koordinaatobjekte leitakse esimeste vaadeldavate objektide seast, võib arvutatavate kauguste arv olla maksimaalne ning olla ülimalt

$$\frac{h(n-h)}{2} .$$

Praktiliselt osutub see arv tunduvalt väiksemaks.

Ka suurust d_{min} pole võimalik täpselt ette anda, sest see eeldaks, et meil oleksid kõigi objektide vahelised kaugused teada. Konkreetse objektimassiivi puhul valime juhuslikult mõned objektid ning leiame nende omavahelised kaugused $d_1 \ldots d_r$. Suuruse d_{min} valime

$$\min \{ d_1, \ldots d_r \} < d_{min} < \max \{ d_1, \ldots d_r \} .$$

Arvestades koordinaatobjektide arvu h_{max} ja kauguse d_{min} omavahelist sõltuvust, on võimalik etteantud d_{min} väärtust ülesande lahendamise käigus parandada, vt. lk. 14.

Olgu antud n objektist koosnev massiiv.

$$O = (O_1 \ldots O_n).$$

Iga objekti iseloomustagu m tunnuse väärtust. Mõõtmistulemused võime siis kirja panna maatrikskujul:

$$A = \begin{Vmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ \cdots & \cdots & \cdots & \cdots \\ a_{i1} & a_{i2} & \cdots & a_{im} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{Vmatrix} ,$$

kus element a_{ij} on i- nda objekti j- nda tunnuse väärtus. Selline objektikogum kujutab endast n punktist koosnevat punktihulka m mõõtmelises ruumis. Objektide omavaheliseks võrdlemiseks kasutatav kaugus võib olla defineeritud mitmesuguselt. Kasutada võib Mahalanobise, Eukleidilisi jt. kaugusi. Käesoleva ülesande lahendamisel kasutame Eukleidilist kaugust. Kahe objekti O_i ja O_j vahel defineerime kauguse :

$$d_{ij} = \sqrt{(a_{i1}-a_{j1})^2 + \ldots + (a_{im}-a_{jm})^2} ,$$

kus suurused a_{ik} ja a_{jk} on nende objektide vastavate tunnuste väärtused. Kui erinevatel tunnustel on muutumispiirkonnad erinevad, siis normeerime tunnuste väärtused:

$$\frac{x_i - \bar{x}_i}{x_{max} - x_{min}} ,$$

kus $\bar{x}_i$, x_{max} ja x_{min} on vastavalt i-nda tunnuse keskmine, maksimaalne ja minimaalne väärtus, vt. /4/.

Esimese koordinaatobjekti valime suvaliselt. Praktiliselt on sobiv selleks valida esimene objekt algandmete maatriksist $KO_1 = O_1$.

(hiljem kasutame tähistust KO_k ka k-nda koordinaatobjekti indeksi tähisena).

Võrdleme objekte alates O_2 -st koordinaatobjektiga KO_1, st. leiame objektide $O_2 \ldots O_k$ kaugused koordinaatobjektist KO_1. Olgu:

1) $d(O_i, KO_1) \leq d_{min}$, kus i= 2...k-1 ,

2) $d(O_i, KO_1) > d_{min}$, kus i= k.

Juhul kui objektid O_i ,kus i= 2... k-1, on küllalt lähedal koordinaatobjektile KO_1 , siis uuteks koordinaatobjektideks neid ei loeta ja nad kuuluvad mittekoordinaatobjektide massiivi $\overline{KO}$

$$O_i \in \overline{KO}$$

Et vähendada tööd kõigile objektidele m lähima koordinaatobjekti leidmisel, fikseerime selle koordinaatobjekti indeksi, mille lähedal antud objekt asus . Selleks defineerime vektori KOI (i) järgmiselt:

$$KOI(i) = \begin{cases} 0, \text{ kui } d(O_i, KO_1) \quad d_{min} , \\ 1, \text{ kui } d(O_i, KO_1) \quad d_{min} . \end{cases}$$

Indeksi i teatud väärtusel k osutub objekt O_k küllalt erinevaks KO_1-st ja ta lisatakse koordinaatobjektide massiivi:

$$KO_2 = O_k .$$

Oletame ,et on leitud s koordinaatobjekti $KO_1 \ldots KO_s$. Nüüd tuleb objektimassiivi objekte O_i võrrelda kõigi olemasolevate koordinaatobjektidega. On kaks võimalust:

1) $d(O_i, KO_s) \leq d_{min}$, kus $s = 2 \ldots h_{max}$,
2) $d(O_i, KO_s) > d_{min}$, mingi s korral.

Juhul, kui objekt O_i asub kas ühe või mitme koordinaatobjekti lähedal, on ta mittekoordinaatobjekt

$$O_i \in \overline{KO} .$$

Sellisel juhul paneme kirja selle esialgse lähima koordinaatobjekti indeksi:

$$\overline{KOI}(i) = \begin{cases} 0, \text{ kui } d(O_i, KO_s) \leq d_{min} , \\ s, \text{ kui } d(O_i, KO_s) > d_{min} . \end{cases}$$

Niipea, kui mingi s väärtusel t objekt O_t asub küllalt kaugel kõigist olemasolevaist koordinaatobjektidest, lisatakse ta koordinaatobjektide massiivi.Objektidemassiivi järgmisi vaatluse alla tulevaid objekte võrreldakse juba ühe objekti võrra suurenenud koordinaatobjektidemassiivi elementidega.

Protseduuri jätkatakse seni, kuni on leitud kõik koordinaatobjektid vastavalt etteantud arvule h_{max} või $h < h_{max}$ koordinaatobjekti.

Võib tekkida olukord, et koordinaatobjektid $KO_1 \ldots KO_{h\ max}$ on kõik leitud, kuid seejuures pole tervet objektimassiivi $O = O_1 \ldots O_n$ läbi vaadatud. Ilmselt on siis ette-antud d_{min} liialt väike, teda tuleks suurendada. Üks võimalus on järgmine.

Oletame, et selleks ajaks millal leiti viimane koordinaatobjekt $KO_{h\ max}$, oli objektimassiivist läbi vaadatud t objekti $O_1 \ldots O_t$. Vaatamata oli jäänud

$$p = \frac{(n-t)}{n} \cdot 100\% \ .$$

Seda arvestades võiks d_{min} väärtust parandada valemiga:

$$d = d_{min} \frac{(100+p)}{100} \ .$$

Seejärel tuleb uut d väärtust kasutades alustada koordinaatobjektide otsimist uuesti algusest, kuni on leitud kõik koordinaatobjektid kogu objektimassiivi ulatuses.

Nagu praktika näitas, suurenes d_{min} sellisel parandamisel liiga palju ja ülesande edasisel lahendamisel ei leitud enam kõiki koordinaatobjekte vastavalt esialgu ette antud arvule h_{max}. Märksa parema tulemuse andis d_{min} parandamine valemiga:

$$d = d_{min} \sqrt{\frac{(100+p)}{100}} \ .$$

Juhul, kui kogu objektimassiiv on läbi vaadatud, kuid leitud koordinaatobjektide arv on väiksem kui h_{max}, siis on ette-antud d_{min} liiga suur. Ülesande lahendamist tuleb alustada uuesti, andes ette väiksema d_{min} väärtuse.

Järgmiseks sammuks on koordinaatobjektide omavaheliste kauguste leidmine, st. leitakse algandmete maatriksist A koordinaatobjektide $KO_1 \dots KO_{h\,max}$ omavahelised kaugused. Saadud tulemused võib esitada maatrikskujul:

$$D_{KO} = \begin{Vmatrix} d_{11} & d_{12} & \dots & d_{1h} \\ \dots & \dots & \dots & \dots \\ d_{k1} & d_{k2} & \dots & d_{kh} \\ \dots & \dots & \dots & \dots \\ d_{h1} & d_{h2} & \dots & d_{hh} \end{Vmatrix} .$$

See on peadiagonaali suhtes sümmeetriline maatriks ($d_{ij} = d_{ji}$) ning peadiagonaalil asuvad nullid.

Järgnevalt leiame iga koordinaatobjekti u lähimat koordinaatobjekti. Selleks leitakse maatriksi D_{KO} igas reas u vähimat elementi:

$$D'_{KO} = \begin{Vmatrix} d'_{11} & d'_{12} & \dots & d'_{1u} \\ \dots & \dots & \dots & \dots \\ d'_{k1} & d'_{k2} & \dots & d'_{ku} \\ \dots & \dots & \dots & \dots \\ d'_{h1} & d'_{h2} & \dots & d'_{hu} \end{Vmatrix} .$$

Iga koordinaatobjekti u lähima koordinaatobjekti indeksid fikseerime maatriksis:

$$KOI = \begin{Vmatrix} KO_{11} & KO_{12} & \dots & KO_{1u} \\ \dots & \dots & \dots & \dots \\ KO_{k1} & KO_{k2} & \dots & KO_{ku} \\ \dots & \dots & \dots & \dots \\ KO_{h1} & KO_{h2} & \dots & KO_{hu} \end{Vmatrix} .$$

Viimaseks sammuks on kõigile objektidele $O_1 \dots O_n$ u lähima koordinaatobjekti leidmine.

Juhul, kui vaadeldav objekt O_i on ise koordinaatobjekt, saame tema lähimate koordinaatobjektide indeksid maatriksist KOI, vastavad kaugused maatriksist D'_{KO}.

Kui objekt O_i ei ole koordinaatobjekt, saame ta lähima koordinaatobjekti indeksi $\overline{KOI}$ -st. Teisi objektile O_i lähimaid koordinaatobjekte otsime tema lähima koordinaatobjekti lähedaste koordinaatobjektide seast (need saame maatriksi KOI vastavast reast). Vaatluse alla võtame kokku objekti O_i (u+1) lähimat koordinaatobjekti, millest leiame u lähimat.

Selleks arvutame algandmete maatriksist kõigi (u+1) lähedase koordinaatobjekti ja antud objekti O_i vahelised kaugused. Neist leiame u vähimat.

Nii leiame kõigi objektide lähimate koordinaatobjektide kaugused ning indeksid:

$$AK = \left\| \begin{matrix} d_{11} & d_{12} & \cdots & d_{1u} \\ \cdots & \cdots & \cdots & \cdots \\ d_{k1} & d_{k2} & \cdots & d_{ku} \\ \cdots & \cdots & \cdots & \cdots \\ d_{n1} & d_{n2} & \cdots & d_{nu} \end{matrix} \right\|$$

$$O = \left\| \begin{matrix} KO_{11} & KO_{12} & \cdots & KO_{1u} \\ \cdots & \cdots & \cdots & \cdots \\ KO_{k1} & KO_{k2} & \cdots & KO_{ku} \\ \cdots & \cdots & \cdots & \cdots \\ KO_{n1} & KO_{n2} & \cdots & KO_{nu} \end{matrix} \right\| .$$

Et edaspidi oleks omavahel sarnaseid objekte mugavam leida, tuleks ka lähimate koordinaatobjektide indeksid järjestada kasvavalt:

$$O' = \begin{Vmatrix} KO'_{11} & KO'_{12} & \ldots KO'_{1u} \\ \ldots & \ldots & \ldots \\ KO'_{k1} & KO'_{k2} & \ldots KO'_{ku} \\ \ldots & \ldots & \ldots \\ KO'_{n1} & KO'_{n2} & \ldots KO'_{nu} \end{Vmatrix}$$

-

VÕHANDU MEETODI PROGRAMM

Programm on koostatud elektronarvutile "Minsk-22". Programm on koostatud automaatses programmeerimiskeeles Malgol. Eeldatud on, et objektimassiiv ei sisalda üle 50 objekti, tunnuste arv on kuni 60 ja ette antav koordinaatobjektide arv on kuni 15 objekti.

Programmi töö käigus trükitakse välja koordinaatobjektide indeksid, iga mittekoordinaatobjekti lähima koordinaatobjekti indeks, koordinaatobjektide kauguste maatriks ning kõigile objektidele vastavalt parameetri TK väärtusele kuni kuue lähima koordinaatobjekti kaugused ja vastavad indeksid.

```
MAGNETIC TAPE 77 PROGRAM 01
COMMENT' TUNNUSPUNKTID' VESKI;
R001 PROCEDURE'KAUGUS(A.,T,S,M,D);BEGIN'
R002 K:=0;
R003 FOR'U:=1 STEP'1 UNTIL'M DO'BEGIN'
R004 K:=K+(A.(S,U).-A.(T,U).)×(A.(S,U).-A.(T,U).);
R005 END';D:=SQRT'(K);END';
R006 ALGUS:READ1'(N,M,HM,DM,TK);
R007 ARRAY'A.(1:N,1:M).,TPC.(1:HM).,TPE.(1:N).;
R010 READARRAY'(A.);K:=0;
R011 FOR'J:=1 STEP'1 UNTIL'M DO'BEGIN'
R012 FOR'I:=1 STEP'1 UNTIL'N DO'BEGIN'
R013 K:=K+A.(I,J).;
R014 K:=K/N;END';
R015 MAX:=A.(1,J).;MIN:=A.(1,J).;
R016 FOR'I:=1 STEP'1 UNTIL'N DO'BEGIN'
R017 IF'A.(I,J).:)MAX THEN'MAX:=A.(I,J).;
R020 IF'A.(I,J).(:MIN THEN'MIN:=A.(I,J).;END';
R021 FOR'I:=1 STEP'1 UNTIL'N DO'
R022 A.(I,J).:=(A.(I,J).-K)/(MAX-MIN);END';
R023 M2:H:=1;TPC.(1).:=1;
R024 FOR'J:=1 STEP'1 UNTIL'N DO'TPE.(J).:=0;
R025 FOR'I:=2 STEP'1 UNTIL'N DO'BEGIN'
R026 FOR'K:=1 STEP'1 UNTIL'H DO'BEGIN'
R027 X:=I;Y:=TPC.(K).;KAUGUS(A.,X,Y,M,D);
R030 IF'D(:DM THEN'GOTO'M;END';
R031 TPC.(H+1).:=I;
R032 IF'H=HM THEN'BEGIN'
R033 DM:=DM×((2×N-I)/N);PRINT1'(DM);GOTO'M2;END';
R034 H:=H+1;GOTO'L;M:TPE.(I).:=TPC.(K).;FOR'R:=K+1 STEP'1 UNTIL'H DO'BEGIN'
R035 KAUGUS(A.,I,TPC.(R).,M,D1);IF'D:)D1 THEN'BEGIN'D:=D1;TPE.(I).:=TPC.(R).;END';END';L:END';HM:=H;
R036 ARRAY'TPI.(1:HM).,TPK.(1:HM,1:HM).,TPLI.(1:HM,1:TK).,TPLK.(1:HM,1:TK).,D1.(1:TK+1).,F1.(1:HM).,F3.(1:TK+1).,B1.(1:N,1:TK).
,B2.(1:N,1:TK).;
R037 FOR'I:=1 STEP'1 UNTIL'HM DO'BEGIN'
R040 TPI.(I).:=TPC.(I).;END';
R041 TEXT10'('KOORDINAATOBJEKTID');OUTPUT'(2);
R042 FOR'I:=1 STEP'1 UNTIL'124 DO'
R043 TEXT10'('=');OUTPUT'(1);K:=0;
R044 FOR'I:=1 STEP'1 UNTIL'HM DO'BEGIN'
R045 TEXTR10'(4,0,TPI.(I).,'I');K:=K+1;
R046 IF'K=20 THEN'BEGIN'K:=0;OUTPUT'(1);END';
R047 END';OUTPUT'(1);
R050 FOR'I:=1 STEP'1 UNTIL'124 DO'
R051 TEXT10'('=');OUTPUT'(3);
R052 TEXT10'('OBJEKTI LAHEDANE KOORDINAATOBJEKT');OUTPUT'(2);
R053 FOR'I:=1 STEP'1 UNTIL'124 DO'
R054 TEXT10'('=');OUTPUT'(1);
R055 FOR'I:=1 STEP'1 UNTIL'N DO'
R056 IF'TPE.(I).=/0 THEN'BEGIN'
R057 TEXTR10'(2,0,'I=',I,':',TPE.(I).,'I');K:=K+1;
R060 IF'K=9 THEN'BEGIN'K:=0;OUTPUT'(1);END';END';
R061 OUTPUT'(1);FOR'I:=1 STEP'1 UNTIL'124 DO'
R062 TEXT10'('=');OUTPUT'(3);
R063 FOR'I:=1 STEP'1 UNTIL'HM DO'BEGIN'
R064 FOR'J:=1 STEP'1 UNTIL'HM DO'BEGIN'
R065 X:=TPI.(I).;Y:=TPI.(J).;KAUGUS(A.,X,Y,M,D);
```

```
TPK. I,J).:=D;END';END';
TEXT10'('KOORDINAATOBJEKTIDE KAUGUSED');OUTPUT'(2);
FOR'I:=1 STEP'1 UNTIL'124 DO'
TEXT10'('=');OUTPUT'(1);
FOR'I:=1 STEP'1 UNTIL'HM DO'BEGIN'
FOR'J:=1 STEP'1 UNTIL'HM DO'
TEXTR10'(2,2,TPK.(I,J).);
OUTPUT'(1);END';
OUTPUT'(1);FOR'I:=1 STEP'1 UNTIL'124 DO'
TEXT10'('=');OUTPUT'(8);
FOR'I:=1 STEP'1 UNTIL'HM DO'BEGIN'
FOR'J:=1 STEP'1 UNTIL'HM DO'BEGIN'
F1.(J).:=0;END';
FOR'V:=1 STEP'1 UNTIL'TK DO'BEGIN'
MIN:=10000;
FOR'J:=1 STEP'1 UNTIL'HM DO'BEGIN'
IF'TPK.(I,J).(;MIN×.I=/J×.F1.(J).=0
 THEN'BEGIN'MIN:=TPK.(I,J).;U:=J;END';END',
F1.(U).:=1;TPLI.(I,V).:=TPI.(U).;TPLK.(I,V).:=TPK.(I,U).;END';END';
TEXT10'('KOLMNURGAD');OUTPUT'(2);
FOR'I:=1 STEP'1 UNTIL'124 DO'
TEXT10'('=');OUTPUT'(1);
FOR'I:=1 STEP'1 UNTIL'HM DO'BEGIN'
FOR'V:=1 STEP'1 UNTIL'TK DO'
TEXTR10'(4,0,TPLI.(I,V).);OUTPUT'(1);
FOR'V:=1 STEP'1 UNTIL'TK DO'
TEXTR10'(2,2,TPLK.(I,V).);OUTPUT'(2);END';
FOR'I:=1 STEP'1 UNTIL'124 DO'
TEXT10'('=');OUTPUT'(8);
                    FOR'I:=1 STEP'1 UNTIL'N DO'BEGIN'
FOR'K:=1 STEP'1 UNTIL'HM DO'
BEGIN'IF'I=TPI.(K). THEN'BEGIN'
FOR'L:=1 STEP'1 UNTIL'TK DO'BEGIN'
B2.(I,L).:=TPLK.(K,L).;B1.(I,L).:=TPLI.(K,L).;
END';GOTO'G;END';END';
X:=I;Y:=TPL.(I).;KAUGUS(A.,X,Y,M,D);
D1.(1).:=D;F3.(1).:=Y;FOR'K:=1 STEP'1 UNTIL'HM DO'IF'Y=TPI.(K). THEN'GOTO'P;
P:FOR'W:=1 STEP'1 UNTIL'TK DO'BEGIN'
X:=I;Y:=TPLI.(K,W).;KAUGUS(A.,X,Y,M,D);
F3.(W+1).:=Y;D1.(W+1).:=D;END';
FOR'WW:=1 STEP'1 UNTIL'TK DO'BEGIN'
FOR'W:=WW STEP'1 UNTIL'TK+1 DO'BEGIN'
IF'D1.(W).(;D1.(WW). THEN'BEGIN'MIN:=D1.(W).;D1.(W).:=D1.(WW).;D1.(WW).:=MIN;
MIN:=F3.(W).;F3.(W).:=F3.(WW).,F3.(WW).:=MIN;END';END';END';
FOR'W:=1 STEP'1 UNTIL'TK DO'BEGIN'B1.(I,W).:=F3.(W).;B2.(I,W).:=D1.(W).;END';
G:END';
TEXT10'('LAHIMAD KOORDINAATOBJEKTID');OUTPUT'(2);
FOR'I:=1 STEP'1 UNTIL'124 DO'
TEXT10'('=');OUTPUT'(1);
FOR'I:=1 STEP'1 UNTIL'N DO'BEGIN'
FOR'V:=1 STEP'1 UNTIL'TK DO'
TEXTR10'(4,0,B1.(I,V).);OUTPUT'(1);END';OUTPUT'(2);
FOR'I:=1 STEP'1 UNTIL'124 DO'
TEXT10'('=');OUTPUT'(8);
TEXT10'('LAHIMATE KOORDINAATOBJEKTIDE KAUGUSED');
OUTPUT'(2);FOR'I:=1 STEP'1 UNTIL'124 DO'
```

```
TEXT10'('=');OUTPUT'(2);
FOR'I:=1 STEP'1 UNTIL'N DO'BEGIN'
FOR'V:=1 STEP'1 UNTIL'TK DO'
TEXTR10'(2,2,B2.(I,V).);OUTPUT'(2);END';
FOR'I:=1 STEP'1 UNTIL'124 DO'
TEXT10'('=');OUTPUT'(3);
STOP';START'ALGUS;FINISH';
```

```
     COMMENT' TUNNUSPUNKTID' VESKI;
MEMORY PLAN
    VARIABLES
  2140  K
  2141  U
  2142  N
  2143  M
  2144  HM
  2145  DM
  2146  TK
  2147  K
  2150  J
  2151  I
  2152  MAX
  2153  MIN
  2154  H
  2155  X
  2156  Y
  2157  D
  2161  R
  2163  D1
  2171  V
  2172  U
  2175  L
  2176  W
  2201  WW
    TABLE OF ARRAYS
  2257  A
  2260  TPC
  2261  TPE
  2262  TPI
  2263  TPK
  2264  TPLI
  2265  TPLK
  2266  D1
  2267  F1
  2270  F3
  2271  B1
  2272  B2
    SUBROUTINES
  2334  2362
  2431  2562
  2443  2451
  2470  2531
  2544  2561
  2602  2606
  2621  2753
  2633  2653
  2723  2752
  3051  3061
  3102  3104
  3122  3144
  3161  3163
  3211  3213
  3230  3264
  3301  3303
  3320  3356
  3332  3355
  3401  3403
  3420  3446
  3432  3443
  3463  3465
  3502  3632
  3514  3520
  3533  3631
  3546  3601
  3652  3654
  3671  3745
  3703  3714
  3731  3742
  3760  3762
  3777  4334
  4011  4061
  4033  4057
  4117  4130
  4143  4173
  4206  4276
  4222  4275
  4311  4333
  4357  4361
  4376  4424
  4410  4421
  4441  4443
  4472  4474
  4511  4537
  4523  4534
  4552  4554
    LABELS
  2370  ALGUS
  2563  M2
  2700  M
  2753  L
  4131  P
  4334  G
    PROCEDURES
  0504  0627  READ1
  0630  0742  ARRAY
  0743  1053  READAR
  1054  1232  PRINT1
  1233  1554  TEXT10
  1555  1564  OUTPUT
  1565  2127  TEXTR1
  2320  2367  KAUGUS
    FUNCTIONS
  0460 - 0503 SQRT
    OPERATING VARIABLES
  2205 - 2210
    CONSTANTS
  2211 - 2256
    PROGRAM
  2307 - 4560
    START 0105
```

N ä i d e.

Objektide massiiv koosneb 24 objektist. Objektideks on antigeenid, vt. /tabel 1/. Tunnuste arv on 24.

Ülesande lahendamisel anti ette suhteliselt suur koordinaatobjektide arv - 12. Kauguse d_{min} ette andmiseks valiti juhuslikult neli objektipaari. Nendevahelised kaugused olid 1,54; 2,42; 1,85; 2,08. Leitud kauguste põhjal ette antud d_{min} = 1,5 osutus väikseks. Suurust d_{min} suurendati 1,7-le. Nagu arvutuste käigus selgus, sobis d_{min}= 1,7 ette antud koordinaatobjektide arvuga 12.

Ülesande esmakordsel lahendamisel leiti igale objektile 3 lähimat koordinaatobjekti. Kui lahenduse kontrollimiseks terve kauguste tabel välja arvutati, selgus et esimese kaheteistkümne objekti seas olevatele mittekoordinaatobjektidele ei olnud leitud lähimaid koordinaatobjekte.

Põhjus oli ilmselt selles, et koordinaatobjektide arv protseduuri alguses on suhteliselt väike. Peale mittekoordinaatobjekti $\bar{O}_j$ lähima koordinaatobjekti KO_i fikseerimist võis mõni hiljem juurde tulnud koordinaatobjekt KO_i' asuda objektile O_j lähemal, kui KO_i.

Kui koordinaatobjekt KO_i' ei asunud KO_i-le küllalt lähedal, st. ei tulnud koordinaatobjekti KO_i kolme lähima koordinaatobjekti hulka, siis mittekoordinaatobjekti O_j lähimate koordinaatobjektide otsimisel jäi KO_i' vaatluse alt hoopis välja.

Siis, kui parameetrile TK anti väärtus 5, st. leiti igale koordinaatobjektile 5 lähimat koordinaatobjekti ja seega ka igale objektile 5 lähimat koordinaatobjekti, muutusid tulemused tunduvalt.

Täieliku kauguste tabeliga võrdlemisel selgus, et saadud tulemused olid kõigi viie lähima koordinaatobjekti jaoks absoluutselt täpsed. Ülesande lahendamiseks kuluv aeg suurenes seejuures väga vähe.

Antigeenid	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1	5	5	5	5	5	3	2	5	4	4	4	4	4	2	3	3	3	2	5	5	4	5	5	3
2	5	4	4	5	5	2	4	5	4	4	5	5	5	2	4	5	4	3	5	5	5	5	5	4
3	3	3	5	3	3	2	2	3	2	3	4	4	4	2	3	3	3	3	4	4	3	5	3	3
4	5	5	3	5	3	2	3	5	4	3	4	5	5	2	3	3	3	3	3	4	3	3	3	3
5	3	4	3	3	4	2	3	5	3	3	4	3	3	2	2	3	2	3	4	4	4	4	4	2
6	4	4	2	4	3	3	3	4	4	4	3	4	3	2	3	3	2	3	3	4	3	4	3	2
7	3	2	2	4	4	2	3	3	3	3	5	3	3	2	2	3	3	3	5	5	4	3	5	3
8	5	5	5	3	5	2	2	5	3	4	5	5	4	2	3	4	3	3	3	5	5	5	5	3
9	5	4	2	3	2	2	3	3	4	2	4	4	3	1	4	3	2	2	2	3	3	3	2	3
10	5	3	3	5	3	2	2	5	5	4	5	5	5	2	3	4	4	4	4	4	5	5	5	3
11	4	5	3	3	3	2	3	5	3	4	5	5	4	2	3	3	3	2	4	4	3	5	5	3
12	5	4	2	5	3	2	3	5	3	3	5	5	3	2	3	3	3	3	3	5	4	4	5	3
13	3	2	2	4	4	3	2	5	4	3	5	5	5	2	3	4	3	3	3	3	3	4	3	3
14	3	4	3	4	4	2	3	5	5	3	3	3	5	3	3	3	3	3	4	3	4	3	3	3
15	3	2	2	4	5	2	4	5	3	2	4	4	5	2	4	3	3	3	4	3	3	2	3	4
16	4	3	2	5	4	3	2	2	3	4	5	3	3	2	3	5	3	2	3	3	3	3	2	3
17	4	3	3	2	5	1	5	5	4	3	4	4	3	2	3	4	3	4	3	4	3	4	2	3
18	4	3	3	5	4	2	3	5	5	4	5	5	5	2	3	2	3	4	3	5	3	5	4	2
19	4	3	3	4	4	2	3	5	4	3	5	4	3	2	3	3	4	5	4	4	4	5	4	2
20	4	3	4	5	3	2	3	3	3	4	5	5	2	2	3	4	3	3	4	5	3	4	2	3
21	3	3	3	5	5	2	2	5	4	5	5	4	3	3	3	4	5	4	5	4	5	5	5	3
22	4	4	4	5	5	3	2	5	5	4	5	5	3	3	4	4	3	3	4	5	3	5	5	3
23	3	3	3	5	5	3	4	5	4	3	5	5	4	3	4	5	4	5	4	5	4	5	5	3
24	3	3	3	5	5	2	3	3	4	3	5	5	5	2	3	4	2	2	5	5	3	5	3	4

Tabel 1

KOORDINAATOBJEKTID

1.I	3.I	4.I	7.I	9.I	10.I	13.I	14.I	16.I	17.I	23.I	24.I

OBJEKTI LAHEDANE KOORDINAATOBJEKT

I= 2.: 1.II= 5.: 3.II= 6.: 4.II= 8.: 1.II= 11.: 4.II= 12.: 4.II= 15.: 13.II= 18.: 4.II= 19.: 10.II= 20.: 3.II= 21.:
10.II= 22.: 1.I

KOORDINAATOBJEKTIDE KAUGUSED

0.00	2.18	1.91	2.38	2.87	1.89	2.53	2.31	2.67	2.64	2.26	2.14
2.18	0.00	2.06	1.98	2.23	2.42	1.91	2.03	2.11	2.01	2.42	1.86
1.91	2.06	0.00	2.49	1.80	1.83	1.85	1.89	2.26	1.99	2.37	2.07
2.38	1.98	2.49	0.00	2.73	2.37	2.23	2.22	2.13	2.30	2.30	1.99
2.87	2.23	1.80	2.73	0.00	2.73	2.24	2.36	1.99	2.11	3.20	2.67
1.89	2.42	1.83	2.37	2.73	0.00	2.01	2.35	2.66	2.54	1.92	2.27
2.53	1.91	1.85	2.23	2.24	2.01	0.00	1.96	1.89	2.11	2.02	1.81
2.31	2.03	1.89	2.22	2.36	2.35	1.96	0.00	2.29	1.99	2.46	2.35
2.67	2.11	2.26	2.13	1.99	2.66	1.89	2.29	0.00	2.42	2.77	2.25
2.64	2.01	1.99	2.30	2.11	2.54	2.11	1.99	2.42	0.00	2.35	2.25
2.26	2.42	2.37	2.30	3.20	1.92	2.02	2.46	2.77	2.35	0.00	2.01
2.14	1.86	2.07	1.99	2.67	2.27	1.81	2.35	2.25	2.25	2.01	0.00

KOLMNURGAD

10.	4.	24.	3.	23.
1.89	1.91	2.14	2.18	2.26
24.	13.	7.	17.	14.
1.86	1.91	1.98	2.01	2.03
9.	10.	13.	14.	1.
1.80	1.83	1.85	1.89	1.91
3.	24.	16.	14.	13.
1.98	1.99	2.13	2.22	2.23
4.	16.	17.	3.	13.
1.80	1.99	2.11	2.23	2.24
4.	1.	23.	13.	24.
1.83	1.89	1.92	2.01	2.27
24.	4.	16.	3.	14.
1.81	1.85	1.89	1.91	1.96
4.	13.	17.	3.	7.
1.89	1.96	1.99	2.03	2.22
13.	9.	3.	7.	24.
1.89	1.99	2.11	2.13	2.25
4.	14.	3.	13.	9.
1.99	1.99	2.01	2.11	2.11
10.	24.	13.	1.	7.
1.92	2.01	2.02	2.26	2.30
13.	3.	7.	23.	4.
1.81	1.86	1.99	2.01	2.07

LAHIMAD KOORDINAATOBJEKTID

10.	4.	24.	3.	23.
10.	1.	23.	24.	4.
24.	13.	7.	17.	14.
9.	10.	13.	14.	1.
7.	14.	3.	17.	13.
4.	9.	14.	13.	1.
3.	24.	16.	14.	13.
1.	10.	4.	23.	3.
4.	16.	17.	3.	13.
4.	1.	23.	13.	24.
4.	1.	13.	10.	9.
4.	10.	1.	13.	9.
24.	4.	16.	3.	14.
4.	13.	17.	3.	7.
13.	14.	4.	24.	3.
13.	9.	3.	7.	24.
4.	14.	3.	13.	9.
4.	10.	13.	1.	14.
10.	23.	13.	4.	1.
3.	24.	17.	7.	13.
23.	10.	1.	13.	24.
23.	1.	24.	10.	4.
10.	24.	13.	1.	7.
13.	3.	7.	23.	4.

LAHIMATE KOORDINAATOBJEKTIDE KAUGUSED

1.89	1.91	2.14	2.18	2.26
1.60	1.64	1.66	2.12	2.24
1.86	1.91	1.98	2.01	2.03
1.80	1.83	1.85	1.89	1.91
1.52	1.61	1.67	1.81	2.07
1.57	1.72	1.76	1.89	2.11
1.98	1.99	2.13	2.22	2.23
1.49	1.71	2.07	2.19	2.23
1.80	1.99	2.11	2.23	2.24
1.83	1.89	1.92	2.01	2.27
1.55	1.78	1.85	1.86	2.22
1.44	1.55	1.85	2.09	2.22
1.81	1.85	1.89	1.91	1.96
1.89	1.96	1.99	2.03	2.22
1.56	1.64	1.99	2.01	2.09
1.89	1.99	2.11	2.13	2.25
1.99	1.99	2.01	2.11	2.11
1.60	1.66	1.79	1.99	2.29
1.53	1.66	1.81	1.89	2.01
1.63	1.68	2.00	2.02	2.07
1.64	1.68	2.09	2.21	2.31
1.47	1.51	1.86	1.89	2.01
1.92	2.01	2.02	2.26	2.30
1.81	1.86	1.99	2.01	2.07

III 1. Järk-järguline rühmitamine.

TA Ajaloo Instituudis tekkis seoses ühe konkreetse tööga vajadus uuritavate objektide rühmitamiseks. Ülesande lahendamiseks sai valitud suhteliselt väikese töömahuga ja lihtsa algoritmiga järk-järguline rühmitamismeetod, vt. /7/.

Meetod kuulub induktiivsete rühmitamismeetodite klassi ja on põhimõtteliselt sarnane Sokal-Michneri meetodiga.

Rühmitamine algab objektidevahelise korrelatsioonimaatriksi leidmisest. Rühmadeks ühendamisel peetakse silmas objektidevahelist maksimaalset korrelatsiooni, rühmade omavahelisel ühendamisel rühmade raskuskeskmete vahelist maksimaalset korrelatsiooni.

Olgu antud n - objekti $O_1 \ldots O_n$, mida soovime rühmitada. Esimesel sammul on rühmade arvuks n. Iga rühm sisaldab vaid üht objekti ning rühmad on tähistatud neisse kuuluvate objektide järgi indeksitega

$$R_1 \ldots R_n .$$

Leiame (n x n) korrelatsioonimaatriksist maksimaalse suuruse r_{ij}, mis esitab objektide O_i ja O_j vahelist korrelatsiooni

$$\max_{i \neq j} r_{ij} = r_{i^* j^*} .$$

Nende kahe objekti kokkuliitmisel moodustatakse uus rühm, mis tähistatakse indeksiga:

$$b = \min \{ i^*, j^* \} .$$

Järgnevalt leitakse antud sammul mitteliitunud objektide ja äsja moodustatud rühma vahelised korrelatsioonid.

Rühma R_b ja objektide O_q vahelised parandatud korrelatsioonid saadakse seostest:

$$r_{bq} = s_{bq} / \sqrt{s_{bb} \cdot s_{qq}} =$$

$$= cov(x_{i^*}+x_{j^*}, x_q)/\sqrt{D(x_{i^*}+x_{j^*}) \cdot D(x_q)} =$$

$$= (s_{i^* q} + s_{j^* q})/\sqrt{(s_{i^* i^*} + s_{j^* j^*} + 2 s_{i^* j^*})} ,$$

kus x_i tähistab indeksile i vastavat objekti ja $s_{i\ q}$ on tavaline objektikogumi kovariatsioon. Saadud korrelatsioonid asendatakse maatriksi b-ndasse ritta ja veergu. Seejärel muudetakse m- s rida ja veerg

$$m = \max \{ i^*, j^* \}$$

mitteoluliseks, asendades vastava rea ja veeru elemendid mingi suure negatiivse arvuga (näit. - 100). Otsime uuesti korrelatsioonimaatriksist suurimat elementi. Teisel sammul vaatleme sisuliselt juba (n-1)x(n-1) -maatriksit, sest esialgsest maatriksist üks rida ja veerg ei tule enam arvesse. Leitud maksimaalse korrelatsiooni põhjal võib sellel sammul ühineda äsja moodustatud rühmaga R_b mingi uus objekt O_i. Suurim korrelatsioon võib kuuluda aga ka mingile uuele objektipaarile ja järgmisel sammul ühendatakse need. Parandused korrelatsioonimaatriksis tehakse analoogiliselt esimesele sammule.

Kolmandal ja järgnevatel sammudel võivad ühineda:

1) kaks rühma (rühmad sisaldavad enam kui üht objekti),

2) kaks objekti (ehk üheobjektilist rühma),

3) rühm ja mingi objekt.

Liites igal sammul 2 rühma, mille raskuskeskmete vaheline korrelatsioon on suurim, vähendatakse rühmade arvu ühe võrra. Samal ajal parandatakse ka korrelatsioonimaatriksit, võttes arvesse muutusi rühmade koosseisus. Rühmitamine lõpetatakse, kui kõik vaadeldavad objektid on ühendatud ühte rühma. Seega suurim rühm sisaldab kõiki objekte. Et rühmitamise alguses oli rühmade arv n, kulus kogu objektimassiivi rühmitamiseks (n-1) sammu. See, millise sammu järel moodustunud rühmad vastavad kõige paremini konkreetsele olukorrale, sõltub lahendatavast ülesandest.

Juhul, kui vaadeldavad objektid on kirjeldatud juba korrelatsioonimaatriksiga, ei ole uute korrelatsioonide arvutamiseks eespool toodud valemeid võimalik kasutada. Sellisel juhul võib minna teist teed: leida moodustunud rühma b ja antud sammul mitteühinenud objektide vahel nn. keskmine korrelatsioon järgmiste valemite abil:

$$z_{i^*q} = \frac{1}{2} \ln \frac{1 + r_{i^*q}}{1 - r_{i^*q}} ,$$

$$z_{j^*q} = \frac{1}{2} \ln \frac{1 + r_{j^*q}}{1 - r_{j^*q}} ,$$

$$\bar{z} = \frac{z_{i^*q} + z_{j^*q}}{2} ,$$

$$\bar{r}_{bq} = \frac{e^{2\bar{z}_{bq}} - 1}{e^{2\bar{z}_{bq}} + 1} .$$

Saadud $\bar{r}_{bq}$ väärtused asendatakse korrelatsioonimaatriksi b- ndasse ritta ja veergu.

Korrelatsioonimaatriksi peadiagonaalil asuvad ühed muudavad logaritmitava avaldise nimetaja juhul, kui $q = i^*$ või $q = j^*$, nulliks, avaldise enda määramatuks. Et aga suurimat r_{ij} otsitakse väljaspool peadiagonaali ja m - s rida ja veerg ($m = \max\{i^*, j^*\}$) muudetakse järgnevalt mitteoluliseks, võib praktilisel arvutamisel asendada peadiagonaali elemendid suvalise arvuga $-1 < a < +1$.

III 2. Järk-järgulise rühmitamismeetodi programm.

Rühmitamise programm on koostatud elektronarvutile "Minsk - 22". Programm on koostatud automaatses programmeerimiskeeles Malgol. Eeldatud on, et objektide massiiv ei ületa 45 objekti ning antud on objektide vaheline samasus-maatriks. Programmi töö käigus trükitakse välja igal sammul ühinenud objektide indeksid.

```
MAGNETIC TAPE 77 PROGRAM 03
COMMENT' CLUST WESKI;
SUBROUTINE'MAX;BEGIN'
M:=R.(1,2).;K:=1;S:=2;
FOR'I:=1 STEP'1 UNTIL'N-1 DO'
FOR'J:=I+1 STEP'1 UNTIL'N DO'BEGIN'
IF'M(:R.(I,J).THEN'BEGIN'
K:=I;S:=J;M:=R.(I,J).;
END';END';
IF'K:)S THEN'BEGIN'
I:=K;K:=S;S:=I;END';
FOR'I:=1 STEP'1 UNTIL'N DO'BEGIN'
IF'R.(I,S).:)-1×.R.(I,K).:)-1 THEN'BEGIN'
Z:=(LN'((1+R.(I,K).)/(1-R.(I,K).))+LN'((1+R.(I,S).)/(1-R.(I,S).)))/2;
Z:=(EXP'(Z)-1)/(EXP'(Z)+1);
R.(I,K).:=Z;R.(K,I).:=Z;
END';END';FOR'I:=1 STEP'1 UNTIL'N DO'BEGIN'R.(S,I).:=-10;R.(I,S).:=-10;END';
FOR'I:=1 STEP'1 UNTIL'N DO'BEGIN'
F.(K,I).:=F.(K,I).+F.(S,I).;
IF'F.(K,I).=/0 THEN'TEXT10'(2,0,I);END';
TEXT10'('

');END';
ALGUS:READ1'(N);OUTPUT'(6);
ARRAY'R.(1:N,1:N).,F.(1:N,1:N).;
READARRAY'(R.);
FOR'I:=1 STEP'1 UNTIL'N DO'
FOR'J:=1 STEP'1 UNTIL'N DO'BEGIN'
IF'I=J THEN'F.(I,J).:=1 ELSE'F.(I,J).:=0;END';
L:=0;FOR'B:=1 STEP'1 UNTIL'N DO'
FOR'W:=B+1 STEP'1 UNTIL'N-1 DO'BEGIN'
IF'R.(B,W).:)-1 THEN'MAX;L:=L+1;IF'L=33 THEN'PRINTA'(R.);END';
STOP';START'ALGUS;FINISH';
```

```
COMMENT' CLUST WESKI;
MEMORY PLAN
  VARIABLES
    1702  M
    1703  K
    1704  S
    1705  I
    1707  N
    1710  J
    1711  Z
    1712  L
    1713  B
    1714  W

  TABLE OF ARRAYS
    1731  R
    1732  F
  SUBROUTINES
    1747  2254  MAX
    1772  2026
    2005  2025
    2050  2154
    2167  2203
    2216  2251
    2311  2344

    2323  2343
    2360  2421
    2375  2420
  LABELS
    2255  ALGUS
  PROCEDURES
    0556  1077  TEXT10
    1100  1223  READ1
    1224  1233  OUTPUT
    1234  1346  ARRAY
    1347  1457  READAR

    1460  1673  PRINTA
  FUNCTIONS
    0516 - 0555 EXP
    0460 - 0515 LN
  OPERATING VARIABLES
    1716 - 1717
  CONSTANTS
    1720 - 1730
  PROGRAM
    1735 - 2423
  START 0105
```

III 3. N ä i d e.

Ülesanne on seotud TA Ajaloo Instituudi sotsioloogia sektoris läbi viidud uurimusega "Teie ja kultuur". Uurimuse üheks eesmärgiks oli uurida raamatukogude lugejaskonna struktuuri, saada raamatukogude töötajaile andmeid kirjanduse soovitamiseks. Lugejail lasti nimetada eesti lemmikkirjanik, lemmikautor vene ja nõukogude kirjanike seast.

Selliselt saadud andmete põhjal oli arvutatud Yule'i korrelatsioonikoefitsient kõigi kirjanike vahel. Tulemused olid esitatud korrelatsioonimaatriksina /vt. tabel 2/.

Vaatluse alla oli võetud 40 sagedamini nimetatud kirjanikku. Esitatud korrelatsioonimaatriks peegeldas kirjanike "lähedust" sõltuvalt teatud lugejagruppide vanusest, haridusest, maitsest jms.

Sobivaiks osutusid rühmad, mis olid moodustunud 34. sammuks. Need rühmad sisaldasid juba enamuse vaadeldavaist kirjanikest. Rühmadesisese struktuuri sai eelnevatel sammudel moodustunud alamrühmi vaadeldes. Kirjanikevaheline "sarnasus" alamrühmade sees oli ilmselt suurem kogu rühma kirjanike omavahelisest sarnasusest. Peale 33. sammu oli korrelatsioonimaatriksi suurimaks elemendiks 0,3503 , korrelatsioon oli tunduvalt vähenenud. Saadud korrelatsioonitabelist sai välja lugeda moodustunud rühmade omavahelised korrelatsioonid, vt. /tabel 3/

Kirjanike loetelu korrelatsioonimaatriksis on järgmine:

1. Koidula
2. Turgenjev
3. Bornhöhe
4. Mälk
5. London
6. Lermontov
7. Jakobson, A.
8. Liiv
9. Balzac
10. Gross
11. Dostojevski
12. Sergo
13. Gorki
14. Kuusberg
15. Metsanurk
16. Jessenin
17. Brontë, C.
18. Sirge
19. Zolà
20. Tagore
21. Simonov
22. Rolland
23. Galsworthy
24. Hemingway
25. Hugo
26. Vaher
27. Sand
28. Puškin
29. Kaugver
30. Tolstoi, A.
31. Tuglas
32. Hint
33. Smuul
34. Dumas
35. Šolohhov
36. Tolstoi, L.
37. Vilde
38. Lacis
39. Luts
40. Tammsaare

Saadud kirjanike rühmitused (kaldkriipsudega on tähistatud alarühmad sarnasuse kohanemise järjekorras):

I / Kuusberg, Kaugver / Brontë, Vaher, Sand, Lacis /

II / Lermontov, Puškin / Koidula, Turgenjev / Bornhöhe, London / Balzac, Hugo, Liiv / /

III / Jessenin, Hemingway / Mälk, Dostojevski, Rolland / Galsworthy, Tolstoi, A., Tammsaare / Tuglas / /

IV / Sirge, Tolstoi, L., / Tagore /

V / Gorki, Vilde / Hint, Smuul /

VI / Jakobson, A., Zolà, Šolohhov / Gross, Metsanurk / Simonov, Dumas, Luts /

Edasine analüüs viis juba nende põhjuste otsimisele, miks teatud kirjanikud osutusid omavahel lähedasteks. Saadud tulemuste põhjal võib lugejaile soovitada uut kirjandust nende lemmikkirjanike järgi.

0,83	0,22	-0,90	0,21	0,64	0,46	0,70	0,09	-0,90	0,07	-0,90	-0,90	-0,90	0,01	-0,04	0,34	-0,01	-0,04	-0,90	-0,11	0,20	0,52	-0,21	0,21	0,20	0,10	-0,69	-0,39	0,18	0,24	0,25	0,18	-0,23	-0,42	0,44	0,15	0,16	0,27	0,27
	0,50	0,49	0,16	0,77	0,42	0,04	0,57	0,39	0,55	0,01	-0,01	0,51	0,62	-0,90	0,30	0,49	0,73	-0,90	-0,90	0,36	0,48	0,11	0,37	0,46	0,63	0,77	0,40	-0,43	0,03	0,07	0,13	-0,28	0,31	0,19	0,49	-0,15	0,29	0,23
		-0,90	0,73	0,62	0,06	0,66	0,56	-0,90	0,54	-0,90	0,71	0,06	-0,06	0,29	0,48	-0,05	0,26	-0,12	-0,17	0,48	-0,39	0,55	0,55	-0,27	0,05	0,49	-0,07	0,28	0,18	0,29	0,22	-0,14	-0,11	0,24	0,48	-0,26	0,02	0,00
			0,47	-0,90	0,57	0,01	0,55	0,37	0,72	0,38	-0,90	-0,90	0,29	0,60	0,42	0,42	0,66	0,43	-0,90	0,74	0,69	0,61	0,45	0,07	0,25	-0,01	0,10	0,73	0,28	0,36	-0,04	0,50	0,53	0,22	0,12	0,25	0,38	0,60
				0,42	0,38	0,37	0,37	0,01	0,35	0,64	0,33	0,29	0,48	0,47	-0,09	-0,90	0,57	0,56	0,52	-0,39	0,31	0,44	0,61	-0,29	-0,90	0,36	0,26	0,60	0,38	0,27	0,37	-0,03	0,27	0,44	0,25	0,25	0,11	0,20
					0,53	0,76	0,62	-0,90	-0,07	-0,90	0,47	0,24	0,24	0,75	0,41	0,22	0,40	0,17	0,12	0,66	0,51	0,24	0,64	0,08	0,35	0,98	0,39	-0,50	0,48	0,21	0,49	-0,36	0,64	0,32	0,38	-0,16	0,40	0,59
						-0,68	-0,68	0,49	0,67	-0,90	0,57	-0,90	0,54	-0,16	-0,90	-0,90	0,57	0,49	0,10	0,25	-0,34	-0,36	-0,01	0,47	-0,05	0,28	-0,21	0,01	-0,59	0,54	0,03	0,33	0,64	0,01	0,31	0,37	0,36	0,49
							0,41	-0,90	0,76	-0,90	0,55	0,52	0,18	0,60	0,50	-0,21	0,13	0,46	0,52	0,47	0,20	0,59	0,52	-0,39	-0,09	0,60	-0,24	0,26	0,42	0,23	0,48	0,04	0,28	0,33	0,09	0,14	-0,05	0,45
								0,26	0,44	0,23	-0,15	-0,90	0,51	0,37	0,16	-0,21	0,14	0,32	0,06	0,48	0,66	0,64	0,73	0,17	0,29	0,53	-0,03	0,44	0,48	0,14	0,24	-0,16	0,38	0,37	0,53	0,08	0,18	0,41
									0,24	0,64	-0,15	0,67	0,72	0,37	0,50	0,32	0,48	0,32	0,41	0,37	0,46	0,33	0,44	0,40	0,13	0,25	0,68	0,44	-0,13	0,49	0,31	0,32	0,56	0,37	0,20	0,44	0,35	0,13
										-0,16	0,19	0,36	-0,21	0,78	0,23	-0,23	0,63	0,13	0,06	0,79	0,52	0,63	0,51	-0,07	-0,43	0,07	-0,05	0,33	0,56	-0,11	0,04	-0,18	0,35	0,20	0,43	-0,06	-0,18	0,52
											0,18	0,14	0,14	-0,23	-0,24	0,33	0,09	-0,28	-0,33	0,33	-0,39	-0,42	0,50	0,29	0,10	0,56	0,19	0,40	0,12	0,11	0,49	-0,00	-0,13	-0,18	0,10	0,47	0,07	0,06
												0,13	0,34	0,57	-0,25	0,63	0,44	0,42	0,34	0,32	-0,68	0,28	0,56	0,12	-0,13	0,55	0,31	0,57	0,11	0,39	0,59	-0,11	0,51	0,17	0,65	0,07	0,33	0,54
													0,30	0,44	0,43	0,43	0,51	0,39	0,53	0,13	0,47	0,45	0,24	0,58	0,49	0,16	0,69	0,35	0,17	0,14	0,14	0,02	0,38	0,44	0,13	0,57	0,24	0,46
														0,29	0,28	0,07	0,69	0,39	0,53	-0,42	0,54	-0,90	0,45	0,45	-0,17	-0,21	0,27	0,62	0,46	0,14	-0,39	0,37	0,55	0,41	0,38	0,60	0,35	0,46
															0,06	0,28	0,25	0,57	0,44	0,65	0,68	0,88	0,52	-0,15	0,31	0,65	0,16	0,42	0,54	-0,07	0,57	-0,26	0,41	0,46	0,16	0,23	-0,16	0,47
																-0,31	0,02	0,67	-0,39	0,11	0,09	-0,47	0,22	0,66	0,67	0,36	0,25	0,25	0,31	0,03	-0,05	0,07	0,36	0,27	0,06	0,71	0,42	0,41
																	0,72	0,22	0,43	0,54	0,36	0,22	0,22	0,34	0,39	-0,02	0,41	0,25	0,24	0,43	0,04	0,07	0,52	0,76	0,20	0,40	-0,02	0,45
																		0,45	0,56	0,70	0,43	0,60	0,32	0,48	-0,22	0,23	0,44	0,54	0,35	0,56	0,22	0,56	0,81	0,31	0,50	0,61	0,25	0,18
																			0,39	0,50	0,20	0,53	0,47	-0,20	0,64	0,21	0,42	0,20	0,33	0,28	0,05	0,41	0,09	0,55	0,35	0,39	0,34	0,46
																				0,38	-0,01	0,54	0,13	-0,55	0,20	-0,32	0,48	0,52	0,28	0,43	0,55	0,59	0,22	048	0,18	0,32	0,41	0,32
																					0,61	0,69	0,59	0,18	0,46	0,47	0,47	0,47	0,42	0,50	0,46	0,42	0,43	0,48	0,20	0,32	0,22	0,54
																						0,35	0,25	0,13	-0,01	0,50	0,29	0,59	0,57	0,13	0,34	-0,13	0,32	0,52	0,14	0,20	0,05	0,58
																							0,25	-0,14	-0,25	0,24	0,27	0,04	0,50	0,17	0,45	0,20	0,45	0,35	0,05	0,14	0,04	0,30
																								0,01	-0,39	0,63	0,21	0,33	0,34	0,31	0,32	-0,04	0,42	0,20	0,37	0,08	0,16	0,47
																									0,55	0,15	0,38	0,20	-0,17	0,27	0,02	0,15	0,41	0,23	0,20	0,29	0,25	0,48
																										-0,11	0,48	0,29	0,02	0,38	0,14	0,45	0,05	0,32	0,05	0,64	0,29	0,52
																											0,11	-0,12	0,38	0,12	0,47	-0,19	-0,01	0,44	0,32	-0,20	0,21	0,20
																												0,26	-0,04	0,09	0,09	0,39	0,36	0,11	0,22	0,47	0,10	0,32
																													0,37	0,28	0,29	0,43	0,58	0,53	0,50	0,26	0,50	0,50
																														0,24	0,37	0,13	0,25	0,31	0,04	0,19	0,20	0,40
																															0,52	0,41	0,39	0,26	0,37	0,53	0,30	0,37
																																0,07	0,49	0,19	0,18	0,29	0,29	0,32
																																	0,43	0,37	0,26	0,54	0,50	0,32
																																		0,27	0,37	0,25	0,35	0,47
																																			0,38	0,31	0,31	0,53
																																				0,35	0,41	0,51
																																					0,47	0,42
																																						0,45

Tabel 2

Rühmad.

6 28
16 24
1 2
19 35
11 22
18 36
4 11 22
3 5
9 25
10 15
1 2 6 28
17 38
14 29
17 27 38
4 11 16 22 24
13 37
7 19 35
17 26 27 38
3 5 8
21 34
23 30
7 10 15 19 35
3 30 40
3 33
3 5 8 9 25
4 11 16 22 23 24 30 40
14 17 26 27 29 38
1 2 3 5 6 8 9 25 28
1 34 39
4 11 16 22 23 24 30 31 40
18 20 36
13 32 33 37
7 10 15 19 21 34 35 39
4 11 16 18 20 22 23 24 30 31 36 40
7 10 13 15 19 21 32 33 34 35 37 39
4 7 10 11 13 15 16 18 19 20 21 22 23 24 30
4 7 10 11 13 14 15 16 17 18 19 20 21 22 23
6 37 38 39 40
4 7 10 11 12 13 14 15 16 17 18 19 20 21 22
5 36 37 38 39 40
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
6 27 28 29 30 31 32 33 34 35 36 37 38 39 40

23	24	30	31	32	33	34	35	36	37	39	40		
21	22	23	24	26	27	29	30	31	32	33	34	35	3
20	21	22	23	24	26	27	29	30	31	32	33	34	3
13	14	15	16	17	18	19	20	21	22	23	24	25	2

39 40

Tabel 3

	I	IV	VII	XIII	XIV	XVIII
I		0,3041	0,0413	0,2599	0,0232	0,0013
IV	0,3041		0,1710	0,2370	0,1113	0,3503
VII	0,0413	0,1710		0,3117	0,2079	0,2767
XIII	0,2599	0,2370	0,3117		0,1501	0,2899
XIV	0,0232	0,1113	0,2079	0,1501		0,3194
XVIII	0,0131	0,35036	0,2767	0,2899	0,3194	

KASUTATUD KIRJANDUS:

1. Lance, G.N., and Williams, W.T. (1966),
   A general theory of classifactory sorting strategies I.
   The Computer Journal, Vol. 9, p. 373.
2. Lance, G.N., Williams, W.T. (1967),
   A general theory of classifactory sorting strategies II.
   The Computer Journal, Vol. 10, p. 241.
3. Frey, T., Võhandu, L.,
   Uus meetod klassifikatsiooniühikute püstitamiseks. ENSV TA Toimetised.
   XV köide, Bioloogiline seeria 1966.
4. Võhandu, L.,
   Kauguse mõiste rakendusi. Matemaatika ja kaasaeg II. Tartu, 1964.
5. Ball, H., What abaut the details?
   AFIPS. Conference proceedings.
   Vol. 27, part. 1. 1965.
6. Bonner, R., On some clustering techniques.
   IBM Journal of. Res. and Devpt., Vol. 8. 1.
7. Sheperd, M., and Willmott, A.,
   Cluster analysis on the Atlas computer.
   The computer Journal, Vol. 2.1.
8. King, B. Step-wise clustering procedures.
   Journal of American Statistical Association, 1967, p. 86 - 101.
9. Goodman, L., and Kruskal, W.,
   Measures of association for cross-classification II,
   Journal American Statistical Association,
   Vol. 54, p. 123.

lo. Sebestyen , G.S.,

Pattern Recognition by an Adaptive Process of Sample Set Construction. IRE Trans. on Info.Theory, Vol. 1 T - 8. 9.

11. Needham, R.M.'

The Theory of Clumps, II. Report M.L. 139, Cambridge Language Research Unit, Cambridge 1961.

12. Sokal, R.R. and Michner, C. D.,

A Quantitative Approach to a Problem in Classification. Evolution , Vol. 11. 6.

13. Hyvarinen, L.,

Classificatoin of Qualitative Data . British Info. Theory J. 1962.

Резюме

В первой главе настоящей дипломной работы дан обзор методов классификаций. Эти методы можно разделить на 2 части:

I индуктивные

П дедуктивные .

Особое внимание уделено индуктивным методам, так как важнейшей чертой индуктивных методов с точки зрения обработки первичных данных является небольшая их трудоемкость по сравнению с дедуктивными методами.

Во второй главе дано описание метода Выханду, дана программа метода и приведен образец. Этот метод уменьшает количество вычислений для классификаций, так как с помощью его находятся кратчайшие расстояния между объектами.

В третьей главе дано описание метода классификаций Б. Кинга, дана программа и приведен образец.

S I S U K O R D:

2. VI 1969. a. A. Veskio