

ԵՐԵՎԱՆԻ ՊԵՏԱԿԱՆ ՀԱՄԱԼՍԱՐԱՆ  
ՄԱԹԵՄԱՏԻԿԱՅԻ ԵՎ ՄԵԽԱՆԻԿԱՅԻ ՖԱԿՈՒԼՏԵՏ  
ԱԿՏՈՒԱՐԱԿԱՆ ԵՎ ՖԻՆԱՆՍԱԿԱՆ ՄԱԹԵՄԱՏԻԿԱ  
ԱՌԿԱ ՈւՍՈՒՑՄԱՆ ԲԱԿԱԼԱՎՐԻ ԾՐԱԳԻՐ

ՀՈՎՍԵՓՅԱՆ ՄԵՐԻ ԷԴԳԱՐԻ

## ԱՎԱՐՏԱԿԱՆ ԱՇԽԱՏԱՆՔ

ՇՈՒԿԱՅԻ ՌԵԺԻՄՆԵՐԻ ՀԱՅՏՆԱԲԵՐՈՒՄ

*« Ֆինանսական մաթեմատիկա » մասնագիտությամբ*

*Ֆինանսական մաթեմատիկայի բակալավրի որակավորման աստիճանի  
հայցման համար*

ԵՐԵՎԱՆ 2023

Ուսանող՝ \_\_\_\_\_  
Հովսեփյան Մերի

Ղեկավար՝ \_\_\_\_\_  
Ֆ.գ.թ., դոցենտ Պողոսյան Միքայել

«Թույլատրելի պաշտպանության»

Ամբիոնի վարիչ՝ \_\_\_\_\_  
տ.գ.թ. դոցենտ Ավետիսյան Կարինե

«\_\_\_\_\_» \_\_\_\_\_ 2023թ

**Շուկայի ռեժիմների հայտնաբերում**

**Выявление рыночный режимов**

**Identification of Market Regimes**

Աշխատանքը նվիրված է շուկայի ռեժիմների հայտնաբերմանը:

Հետազոտության հիմնական նպատակն է բացահայտել շուկայական ռեժիմները, հասկանալ ֆինանսական շուկաների հիմքում ընկած օրինաչափությունները և վարքագիծը: Այս հետազոտության արդյունքները պատկերացումներ են տալիս շուկայական ռեժիմների հայտնաբերման վերաբերյալ, որոնք կարող են օգտագործվել ներդրողների և պորտֆելի կառավարիչների կողմից ռիսկերի կառավարման, ակտիվների բաշխման և ներդրումային ռազմավարությունների բարելավման համար:

Հիմնական խնդիրը ֆինանսական շուկաներում հստակ շուկայական ռեժիմների նույնականացումն ու բնութագրումն է՝ օգտագործելով մեքենայական ուսուցման կլաստերավորման տեխնիկան և Թաքնված Մարկովյան Մոդելները:

**Բանալի բառեր՝** Կլաստերավորում, շուկայական ռեժիմներ, ներդրումային որոշումներ, Թաքնված Մարկովյան Մոդել:

## ԲՈՎԱՆԴԱԿՈՒԹՅՈՒՆ

ՆԵՐԱԾՈՒԹՅՈՒՆ.....	5
ԳԼՈՒԽ 1. ՇՈՒԿԱՅԱԿԱՆ ՌԵԺԻՄՆԵՐ, ԳՐԱԿԱՆՈՒԹՅԱՆ ԱԿՆԱՐԿ .....	7
1.1. Շուկայական ռեժիմներ .....	7
1.2. Գրականության ակնարկ .....	10
ԳԼՈՒԽ 2. ՄԵՔԵՆԱՅԱԿԱՆ ՈՒՍՈՑՈՒՄ, ՄԵԹՈԴՆԵՐ .....	11
2.1 Մեքենայական ուսուցում .....	11
2.1.1 Չվերահսկվող ուսուցում .....	13
2.2 K-means ալգորիթմ .....	17
2.3 Գաուսյան Խառնուրդի Մոդել .....	19
ԳԼՈՒԽ 3. ԹԱՔՆՎԱԾ ՄԱՐԿՈՎՅԱՆ ՄՈԴԵԼՆԵՐ .....	21
3.1 Մարկովյան շղթա .....	21
3.2 Թաքնված Մարկովյան Մոդելներ .....	23
3.2.1 Թաքնված Մարկովյան Մոդելի Ուսուցում .....	25
3.2.2 Թաքնված Մարկովյան Մոդելի Վերծանում .....	29
ԳԼՈՒԽ 4. ՎԵՐԼՈՒԾՈՒԹՅՈՒՆ.....	31
ԵԶՐԱԿԱՑՈՒԹՅՈՒՆ .....	45
ՕԳՏԱԳՈՐԾՎԱԾ ԳՐԱԿԱՆՈՒԹՅԱՆ ՑԱՆԿ .....	46

## ՆԵՐԱԾՈՒԹՅՈՒՆ

Շուկան փոխում է իր վարքագիծը ժամանակի ընթացքում, ինչը հանգեցնում է շուկայական տրամադրությունների փոփոխությանը:

Ֆինանսական աշխարհում տարբեր գործոններ ազդում են շուկաների վրա, և այդ ազդեցությունը բնութագրելը կարևոր է ներդրողների, պորտֆելի կառավարիչների և ներգրավված այլ անձանց համար՝ հասկանալու այս շուկաների օրինաչափությունները և դինամիկան:

Նախքան ֆինանսական շուկայի մասնակիցները կսկսեն փնտրել եկամտաբեր առևտրային հնարավորություններ, կարևոր է, որ դիտարկեն շուկայի ներկա իրավիճակը և փորձեն կանխատեսել շուկայի հետագա վարքագիծը, որպեսզի կարողանան որոշել համապատասխան առևտրային ռազմավարություններ:

Բացահայտելով շուկայական ռեժիմները և հասկանալով դրանց անցումները՝ մասնագետները կարող են ավելի լավ տեղեկացված որոշումներ կայացնել, օպտիմալացնել իրենց ներդրումային ռազմավարությունները և ավելի արդյունավետ կառավարել ռիսկերը:

Աշխատանքի արդիականությանը նպաստում է մեքենայական ուսուցման ալգորիթմների և տեխնիկայի արագ առաջընթացը, որի շնորհիվ զգալիորեն մեծացել է բարդ ֆինանսական տվյալները մշակելու, վերլուծելու կարողությունը, ինչպես նաև առաջադեմ ալգորիթմների կիրառումը, ինչպիսին է Թաքնված Մարկովյան Մոդելը:

**Հետազոտության նպատակը և խնդիրները:** Հետազոտության հիմնական նպատակն է բացահայտել և վերլուծել շուկայական ռեժիմները ֆինանսական շուկաներում՝ օգտագործելով մեքենայական ուսուցման կլաստերավորման տեխնիկան և Թաքնված Մարկովյան Մոդելները:

Ելնելով առաջադրված նպատակներից՝ դրվել են հետևյալ խնդիրները.

1. Կլաստերավորման ալգորիթմի ընտրության խնդիր: Համեմատել և գնահատել մեքենայական ուսուցման տարբեր կլաստերավորման մեթոդներ՝ շուկայական ռեժիմների արդյունավետ հայտնաբերման և բնութագրման ամենահարմար մեթոդը որոշելու համար:

2. Թաքնված Մարկովյան Մոդելի մշակման խնդիր: Համապատասխան Թաքնված Մարկովյան Մոդելի կառուցում, որը ճշգրիտ կերպով արտացոլում է ռեժիմի փոփոխությունների հավանականային բնույթը և մոդելավորում է անցումները հայտնաբերված շուկայական ռեժիմների միջև:

**Հետազոտության օբյեկտը և առարկան:** Հետազոտության օբյեկտ է հանդիսանում ֆինանսական շուկան, մասնավորապես՝ S&P 500 ինդեքսը, իսկ առարկան վերաբերվում է առաջադեմ հաշվողական մեթոդների կիրառմանը, ինչպիսիք են մեքենայական ուսուցման կլաստերավորման ալգորիթմները և Թաքնված Մարկովյան Մոդելները:

**Հետազոտության արդյունքների կիրառական նշանակությունը:**

Հետազոտության արդյունքները կիրառական նշանակություն ունեն ներդրողներին, պորտֆելի կառավարիչներին և վերլուծաբաններին օգնելու ավելի լավ հասկանալ շուկայի դինամիկան, կառավարել ռիսկերը, օպտիմալացնել ակտիվների բաշխումը, ժամանակին ներդրումային որոշումներ կայացնել և կատարելագործել առևտրային ռազմավարությունները:

**Ավարտական աշխատանքի կառուցվածքը:** Աշխատանքը բաղկացած է 4 գլուխներից: Առաջին գլխում ներկայացված են շուկայական ռեժիմները և տրված է գրականության ակնարկ: Երկրորդ գլուխը նվիրված է մեքենայական ուսուցմանը և նրա ալգորիթմներին: Երրորդ գլխում ներկայացված է Թաքնված Մարկովյան Մոդելները և այն խնդիրները որոնք լուծվում են Թաքնված Մարկովյան Մոդելների միջոցով: Աշխատանքի չորրորդ գլխում իրականացվել է վերլուծական աշխատանք՝ հիմք ընդունելով S&P 500-ի և VIX-ի ժամանակային շարքի տվյալները:

## ԳԼՈՒԽ 1. ՇՈՒԿԱՑԱԿԱՆ ՌԵԺԻՄՆԵՐ, ԳՐԱԿԱՆՈՒԹՅԱՆ ԱԿՆԱՐԿ

### 1.1. Շուկայական ռեժիմներ

Ֆինանսական շուկաները համաշխարհային տնտեսության կենսական մասն են՝ ծառայելով որպես հարթակ, որտեղ անհատները, ընկերությունները և կառավարությունները կարող են առևտուր անել ֆինանսական ակտիվներով, ինչպիսիք են բաժնետոմսերը, պարտատոմսերը, արժույթները և ապրանքները:

Ֆինանսական շուկաները ներառում են տարբեր հարթակներ և համակարգեր, ներառյալ ֆոնդային շուկաները, պարտատոմսերի շուկաները, ապրանքային շուկաները և արտարժույթի շուկաները, որոնք հեշտացնում են ֆինանսական գործիքների առևտուրը:

Ֆինանսական շուկաները ժամանակի ընթացքում ենթարկվում են տարբեր փոփոխությունների, որոնք բնութագրվում են շուկայական տարբեր ռեժիմներով:

Ֆինանսական շուկաների դինամիկան հասկանալի, ներառյալ շուկայական ռեժիմները և շուկայի վարքագծի վրա ազդող այլ գործոններ, կարևոր է շուկայի մասնակիցների համար տեղեկացված ներդրումային որոշումներ կայացնելիս և ռիսկերն արդյունավետ կառավարելիս:

Շուկայական ռեժիմները վերաբերում են ֆոնդային շուկայի տարբեր փուլերին, որոնք բնութագրվում են շուկայական վարքագծի հստակ ձևերով, ինչպիսիք են բաժնետոմսերի գները, տատանողականությունը և առևտրի ծավալը:

Շուկայական ռեժիմները դասակարգվում են՝ ելնելով ակտիվների եկամտաբերության, տատանողականության և ակտիվների միջև հարաբերակցության օրինաչափություններից:

Շուկայական ռեժիմների տեսակներն են՝

- Ցուլի շուկա (Bull market)
- Արջի շուկա (Bear market)
- Չեզոք շուկա (Neutral market)

Ցուլի շուկան ուղեկցվում է ակտիվների գների կայուն աճով, որը բնութագրվում է ներդրողների լավատեսությամբ, ռիսկի բարձրացման և շուկայական ինդեքսների

ընդհանուր աճի միտումով: Նման ժամանակներում ներդրողները հաճախ հավատում են, որ աճի միտումը կշարունակվի երկարաժամկետ հեռանկարում: Ներդրողները պատրաստ են ստանձնել բարձր ռիսկեր, ավելի լավ եկամտաբերություն ունենալու համար, իսկ ընդհանուր շուկան, որը չափվում է շուկայական ինդեքսներով, ցուցադրում է ակտիվների գների հետևողական աճ:

Արջի շուկան ակտիվների գների երկարատև անկման շրջան է, որը բնութագրվում է ներդրողների հոռետեսությամբ, ռիսկից խուսափումով և շուկայական ինդեքսների ընդհանուր նվազման միտումով: Արջի շուկայի ժամանակ տնտեսության գործունեությունը դանդաղում է, և գործազրկությունն աճում է, քանի որ ընկերությունները սկսում են աշխատողներին կրճատել: Արջի շուկայի ժամանակ ներդրողները կարող են իրենց ուշադրությունը կենտրոնացնել դեպի ավելի ապահով ակտիվներ, ինչպիսիք են պարտատոմսերը կամ կանխիկ միջոցները, որպեսզի պաշտպանեն իրենց պորտֆելները հետագա անկումից:

Չեզոք շուկան չի ցուցաբերում հստակ աճի կամ նվազման միտում, և ակտիվների գները մնում են համեմատաբար կայուն որոշակի միջակայքում: Շուկայի ընդհանուր ուղղությունը այս համատեքստում մնում է անորոշ:

Տարբեր գործոններ կարող են ազդել շուկայական ռեժիմների փոփոխության վրա, ինչպիսիք են՝

- Տնտեսական պայմաններ
- Դրամավարկային քաղաքականություն
- Շուկայական տրամադրություններ
- Աշխարհաքաղաքական իրադարձություններ
- Տեխնոլոգիական առաջընթաց և այլն

Տնտեսության ընդհանուր վիճակը, ներառյալ ՀՆԱ-ի աճը, գնաճը, զբաղվածության մակարդակը, կարող են ազդել շուկայական ռեժիմների վրա: Օրինակ, ցածր գնաճով և զբաղվածության բարձր մակարդակով ուժեղ տնտեսությունը կարող է հանգեցնել շուկայի աճի, մինչդեռ տնտեսական անկումը կարող է հակառակ ազդեցությունը ունենալ:



Կենտրոնական բանկերի որոշումները տոկոսադրույքների և դրամավարկային քաղաքականության այլ գործիքների վերաբերյալ կարող են էական ազդեցություն ունենալ ֆինանսական շուկաների վրա: Ցածր տոկոսադրույքները կարող են խթանել տնտեսական աճը և հանգեցնել շուկայի աճի, մինչդեռ բարձր տոկոսադրույքները կարող են դանդաղեցնել աճը:

Ներդրողների տրամադրությունները կամ շուկայի մասնակիցների ընդհանուր վերաբերմունքը ֆինանսական շուկաների նկատմամբ կարող է էական դեր ունենալ շուկայական ռեժիմների որոշման գործում: Լավատեսությունն ու վստահությունը կարող է նպաստել շուկայի աճին, մինչդեռ հոռետեսությունն ու վախը կարող է նպաստել արջի շուկայի զարգացմանը:

Իրադարձությունները, ինչպիսիք են՝ պատերազմները, քաղաքական անկայունությունը կամ առևտրային վեճերը, կարող են անորոշություն և ռիսկ առաջացնել ֆինանսական շուկաներում՝ պոտենցիալ փոփոխություններ առաջացնելով շուկայական ռեժիմներում:

Տեխնոլոգիաների առաջընթացը կարող է ստեղծել ներդրումային հնարավորություններ և խթանել շուկայի աճը: Ներդրումների աճը կարող է բարձրացնել ակտիվների գները և նպաստել շուկայի ընդհանուր աճին, ինչը կարող է հանգեցնել շուկայի աճի՝ ներդրողների վստահության և շուկայի դրական տրամադրությունների արդյունքում:

Շուկայական ռեժիմների բացահայտումը կարևոր է ներդրողների, առևտրականների և պորտֆելի կառավարիչների համար, քանի որ այն թույլ է տալիս նրանց հարմարեցնել իրենց ռազմավարությունները և ռիսկերի կառավարման տեխնիկան՝ համաձայն շուկայական գերիշխող վարքագծի: Հայտնաբերելով ներկայիս շուկայական ռեժիմը՝ շուկայի մասնակիցները կարող են օպտիմալացնել իրենց ակտիվների բաշխումները, բարելավել ռիսկերի կառավարումը և բարձրացնել իրենց պորտֆելի ընդհանուր կատարողականը: Տարբեր տեխնիկաներ, ինչպիսիք են մեքենայական ուսուցման կլաստերավորման մեթոդները և Թաքնված Մարկովյան Մոդելները, կարող են օգտագործվել շուկայական ռեժիմները բացահայտելու և ռեժիմի փոփոխությունները կանխատեսելու համար:

## 1.2. Գրականության ակնարկ

Տարիների ընթացքում տարբեր հետազոտողներ, վերլուծաբաններ, տնտեսագետներ փորձել են հասկանալ շուկայի դինամիկան և կանխատեսել հնարավոր փոփոխությունները:

Որոշ կայքեր, կրթական հարթակներ կատարել են հետազոտություն շուկայի ռեժիմների հայտնաբերման վերաբերյալ, դրանցից է՝ QuantStart<sup>1</sup> կրթական հարթակը, որը բացատրում է, թե ինչպես օգտագործել թաքնված Մարկովյան Մոդելները ֆինանսական ժամանակային շարքերի տվյալները վերլուծելու և շուկայական ռեժիմների փոփոխությունները բացահայտելու համար, R- ծրագրավորման լեզվի միջոցով:

Հաջորդ աշխատությունը՝ Շուկայական ռեժիմի հայտնաբերումն է, օգտագործելով վիճակագրական և Մեքենայական ուսուցման վրա հիմնված մոտեցումները<sup>2</sup>, հեղինակներն են՝ Հայկազ Արամյանը, Ջեյսոն Ռամչանդանին և Մարիոս Սքեվոֆիլակասը: Այս հոդվածի նպատակն է պարզել S&P 500-ի նորմալ (աճի) կամ կտրուկ անկման շուկայական վիճակները՝ օգտագործելով մի քանի վիճակագրական և մեքենայական ուսուցման մոդելներ, ինչպիսիք են՝ Գաուսյան Թաքնված Մարկովյան Մոդելները, k-means կլաստերավորումը և Գաուսյան խառնուրդների մոդելները: Այս նախագծում կիրառել են S&P 500-ի վրա ֆյուչերսային պայմանագրերը: Բացի այդ, կառուցել են պարզ ներդրումային ռազմավարություններ՝ հիմնվելով բացահայտված շուկայական ռեժիմների վրա:

## ԳԼՈՒԽ 2. ՄԵՔԵՆԱՅԱԿԱՆ ՈՒՍՈՒՑՈՒՄ, ՄԵԹՈԴՆԵՐ

### 2.1 Մեքենայական ուսուցում

Մենք գտնվում ենք մի դարաշրջանում, որտեղ առկա է ահռելի թվով տվյալների ներհոսք, և այս տվյալների հեղեղը պահանջում է տվյալների ավտոմատացված վերլուծություն: Պատկերացրեք, որ մենք նավարկում ենք տվյալների անսահման օվկիանոսում, մեքենայական ուսուցումը ծառայում է որպես կողմնացույց, որն ուղղորդում է մեր որոնումները՝ զինելով մեզ ավտոմատացված գործիքներով՝ վերլուծելու, վերծանելու և ներսում թաքնված գիտելիքները օգտագործելու համար:

Մեքենայական ուսուցումը<sup>3</sup> (Machine Learning-ML) արհեստական բանականության (Artificial Intelligence-AI) տեսակ է, որը թույլ է տալիս ծրագրային ապահովման հավելվածներին ավելի ճշգրիտ լինել արդյունքները կանխատեսելիս՝ առանց դրա համար հստակ ծրագրավորված լինելու:

1959 թվականին Արթուր Սամուելը սահմանեց մեքենայական ուսուցումը այսպես՝ «Հետազոտության ոլորտ է, որը հնարավորություն է տալիս մեքենային սովորել այն, ինչ բացահայտ ծրագրավորած չէ»:

Մասնավորապես մեքենայական ուսուցումը սահմանվում է, որպես մեթոդների մի շարք որոնք կարող են ավտոմատ կերպով հայտնաբերել օրինաչափությունները տվյալների մեջ:

Մեքենայական ուսուցումը բաժանվում է երեք հիմնական տեսակի՝

- Վերահսկվող ուսուցում (Supervised learning)
- Առանց վերահսկման ուսուցում (Unsupervised learning)
- Ամրապնդմամբ ուսուցում (Reinforcement learning )

Վերահսկվող ուսուցման ժամանակ ալգորիթմը վերապատրաստվում է պիտակավորված տվյալների բազայի հիման վրա, որտեղ հայտնի են մուտքային-ելքային տվյալների կապը: Երբ ասում ենք՝ պիտակավորված տվյալներ, նկատի ունենք մի տվյալների հավաքածու, որը ներառում է մուտքային-ելքային զույգեր, որտեղ յուրաքանչյուր մուտք ունի իրեն համապատասխան ելքային արժեք, որոնք կոչվում են՝

պիտակներ: Այս պիտակները անհրաժեշտ են հիմքում ընկած օրինաչափությունները կամ մուտքերի և արդյունքների միջև փոխհարաբերությունները սովորելու համար:

Պիտակավորված տվյալների օրինակ է տվյալների բազան, որն օգտագործվում է էլեկտրոնային նամակների դասակարգիչ ստեղծելու համար, այն է որ, ստացված հաղորդագրությունները կեղծ(spam) են թե ոչ կեղծ(ham): Այս դեպքում տվյալների բազան բաղկացած է էլեկտրոնային նամակների որոշ նմուշներից, որոնք մուտքերն են, դրանց համապատասխան դասակարգումներից(ելքերից)՝ որպես «spam» կամ «ham»:

Ստորև աղյուսակում ներկայացված է պիտակավորված տվյալների օրինակ:

Աղյուսակ 1: Պիտակավորված տվյալների օրինակ

Էլեկտրոնային նամակներ	Պիտակներ
«Շնորհավորում եմ, դուք շահել եք անվճար ճանապարհորդություն»:	Կեղծ
Հանդիպում կիրակի օրվա համար:	Ոչ կեղծ
«Ձեր պատվերը առաքվել է»:	Ոչ կեղծ
«Դուք շահել եք 1.000.000 դրամ, անցեք հղումով գումարը ստանալու համար»	Կեղծ

Չվերահսկվող ուսուցման ժամանակ ալգորիթմը վերապատրաստվում է չպիտակավորված տվյալների բազայի հիման վրա, և նպատակը տվյալների ներսում թաքնված օրինաչափությունները հայտնաբերելն է:

Ամրապնդմամբ ուսուցումը մեքենայական ուսուցման մեթոդի մի տեսակ է, որտեղ գործակալը (համակարգչային ծրագիրը) փոխադրում է շրջակա միջավայրի հետ և սովորում է գործել դրա շրջանակներում:

Այս աշխատանքի շրջանակներում անդրադառնալու ենք չվերահսկվող ուսուցմանը, քանի որ շուկայական ռեժիմների նույնականացումը չվերահսկվող գործընթաց է: Օգտագործելով չվերահսկվող ուսուցման տեխնիկան, մենք նպատակ ունենք բացահայտել թաքնված օրինաչափություններ և կառուցվածքները ֆինանսական տվյալների մեջ՝ թույլ տալով մեզ արդյունավետորեն բացահայտել

շուկայական տարբեր ռեժիմներ՝ առանց հենվելու նախապես սահմանված պիտակների կամ նախնական գիտելիքների վրա:

### 2.1.1 Չվերահսկվող ուսուցում

Չվերահսկվող ուսուցման ժամանակ մեզ պարզապես տրվում են մուտքային տվյալներ, առանց որևէ պիտակի:

Չվերահսկվող ուսուցման ավգորիթմներն օգտագործում են չպիտակավորված տվյալներ՝ տվյալների ներսում թաքնված կառուցվածքները, օրինաչափությունները կամ հարաբերությունները բացահայտելու համար:

Չպիտակավորված տվյալների հայտնի օրինակ է տվյալների բազան, որը պարունակում է առցանց խանութի հաճախորդների գործարքների մասին տեղեկատվություն: Տվյալների հավաքածուն կարող է ներառել ատրիբուտներ, ինչպիսիք են հաճախորդի ID-ն, տարիքը, սեռը և գնումների պատմությունը, սակայն չկան պիտակներ, որոնք ցույց են տալիս նախապես սահմանված կատեգորիաներ կամ խմբեր: Ստորև աղյուսակում ներկայացված է հետևյալ չպիտակավորված տվյալների օրինակը:

Աղյուսակ 2: Չպիտակավորված տվյալների օրինակ

Հաճախորդի ID	Տարիք	Սեռ	Գնումների պատմություն
1	20	Աղջիկ	{Վերնաշապիկ, խնամքի պարագաներ, հեծանիվ}
2	37	Տղա	{Հեռախոս, համակարգիչ, մեքենա}
3	25	Տղա	{Կոշիկ, ականջակալ, սպորտային ժամացույց}
4	41	Աղջիկ	{Զգեստ, սանդալներ, վզնոց}

Այս դեպքում յուրաքանչյուր հաճախորդի հետ կապված կոնկրետ ելքային պիտակներ չկան: Չվերահսկվող ուսուցման ավգորիթմը, ինչպիսին է

կլաստերավորումը, կարող է կիրառվել այս տվյալների բազայի վրա՝ խմբավորելու հաճախորդներին՝ հիմնվելով տարիքի, սեռի և գնումների պատմության նրանց նմանությունների վրա: Մա կարող է օգնել խանութին բացահայտել հաճախորդների սեգմենտները նպատակային մարքեթինգային արշավների կամ արտադրանքի առաջարկությունների համար՝ առանց նախապես սահմանված կատեգորիաների վերաբերյալ որևէ նախնական տեղեկություն պահանջելու:

Չվերահսկվող ուսուցման երկու հիմնական խնդիրներից են՝

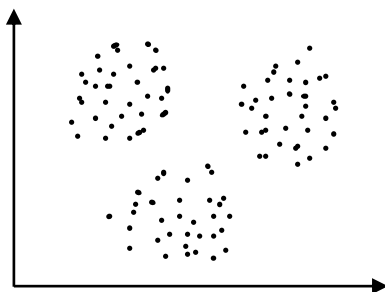
- Կլաստերավորում
- Չափողականության նվազեցում

Տեխնիկական տեսակետից՝ չափողականության նվազեցումը տվյալների բարդության նվազեցման գործընթաց է՝ որոշակի չափով պահպանելով դրա կառուցվածքի համապատասխան մասերը:

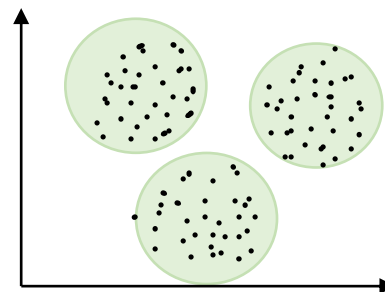
Աշխատանքի շրջանակներում անդրադառնալու ենք կլաստերավորմանը և նրա մեթոդներին:

Կլաստերային վերլուծությունը կամ կլաստերավորումը<sup>4</sup> օբյեկտների հավաքածուն խմբավորելու խնդիրն է այնպես, որ նույն խմբի օբյեկտները (կոչվում են կլաստերներ) ավելի նման լինեն (որոշ իմաստով) միմյանց, քան մյուս խմբերի (կլաստերի) օբյեկտներին: Համակարգը մեր կողմից մուտքագրած տվյալները, փորձում է ինչ-որ ընդհանուր հատկություններից ելնելով բաժանել խմբերի, այլ կերպ ասած կլաստերների:

Գծապատկեր 1: Կլաստերավորումից առաջ



Գծապատկեր 2: Կլաստերավորումից հետո



Վերը նշված գծապատկերներում հավաքված տվյալների կետերը կարող են դասակարգվել երեք տարբեր խմբերի(կլաստերների), ինչն էլ արտացոլվում է Գծապատկեր 2-ում:

Կլաստերավորման<sup>5</sup> մեջ առաջնային խնդիրը կլաստերների քանակի ընտրությունն է: Նշանակենք  $K$ -ով կլաստերների թիվը: Առաջնային նպատակն է գնահատել կլաստերների քանակի բաշխումը՝  $P(K|D)$ , որտեղ  $D$ -ն տվյալների հավաքածուն է: Չվերահսկվող ուսուցման և կլաստերավորման համատեքստում այն ներկայացնում է տվյալների մեջ որոշակի քանակությամբ կլաստերներ ունենալու հավանականությունը:

Չվերահսկվող ուսուցման ժամանակ մենք ազատ ենք ընտրելու այնքան կլաստերներ, որքան ցանկանում ենք, բայց նպատակն է գտնել կլաստերների օպտիմալ թիվը, որը լավագույնս ներկայացնում է տվյալների հիմքում ընկած կառուցվածքը:

Դա անելու համար մենք գնահատում ենք  $P(K|D)$  բաշխումը և գտնում  $K$ -ի արժեքը, որը առավելագույնի է հասցնում այս հավանականությունը: Պարզության համար մենք հաճախ գնահատում ենք այն՝

$$K^* = \operatorname{argmax}_k p(K|D):$$

Այլ կերպ ասած վերը նշված բանաձևը նշանակում է գտնել այն  $K$ -ի արժեքը, որը առավելագույնի է հասցնում  $P(K|D)$  հավանականության բաշխումը: Ըստ էության,  $K^*$ -ը կլաստերների օպտիմալ թիվն է, որը լավագույնս համապատասխանում է տրված տվյալներին:

Երկրորդ խնդիրն է գնահատել, թե յուրաքանչյուր կետ որ կլաստերին է պատկանում: Եթե սա պրոեկտենք մեր աշխատանքի վրա, ապա նպատակը կլինի որոշել, թե շուկայական որ ռեժիմին է պատկանում տվյալների յուրաքանչյուր կետը:

Թող  $z_i \in \{1, \dots, K\}$  ներկայացնի այն կլաստերը, որին վերագրված է  $i$ -տվյալների կետը: Մենք կարող ենք եզրակացնել թե որ կլաստերին է պատկանում յուրաքանչյուր տվյալների կետ՝ հաշվարկելով՝

$$z_i = \operatorname{argmax}_k p(z_i = k|x_i, D):$$

Մենք ցանկանում ենք յուրաքանչյուր  $x_i$  տվյալների կետին վերագրել կլաստերային պիտակ՝  $z_i$ , այլ կերպ ասած, յուրաքանչյուր կետ պատկանում է  $k$  կլաստերներից մեկին:  $p(z_i = k|x_i, D)$ -ն ներկայացնում է  $k$ -կլաստերին պատկանող  $x_i$  կետի հավանականությունը, հաշվի առնելով  $D$ -տվյալների բազմությունը: Այնուհետև

մաքսիմալացումը օգտագործվում է յուրաքանչյուր  $x_i$  կլաստերի պիտակը գնահատելու համար, ինչը նշանակում է, որ մենք ընտրում ենք  $k$ -կլաստերի պիտակը, որը առավելագույնի է հասցնում  $x_i$ -ի  $k$ -ին պատկանելու հավանականությունը:

Հաճախորդների սեգմենտավորման օրինակում, մենք ցանկանում ենք խմբավորել գնումները  $K$  խմբերում՝ հասկանալու հաճախորդների վարքագիծը:  $K$  կլաստերներ ունենալուց հետո, մենք կարող ենք գնահատել թե որ կլաստերին է պատկանում յուրաքանչյուր գնում՝ հաշվարկելով  $K$  կլաստերներին պատկանող յուրաքանչյուր գնման հավանականությունը՝ հաշվի առնելով տվյալների հավթաժուն: Այնուհետև մենք յուրաքանչյուր գնում վերագրում ենք ամենաբարձր հավանականությամբ կլաստերին: Սա կարող է օգնել մեզ հասկանալ, թե որ ապրանքներն են հաճախ գնում միասին, որ հաճախորդներն են ամենայն հավանականությամբ գնելու որոշակի ապրանքներ և այլ պատկերացումներ, որոնք կարելի է քաղել կլաստերներից:

Կլաստերավորման ալգորիթմների տեսակներն են՝

- Հիերարխիկ կլաստերավորում(Hierarchical clustering)-հիերարխիկ կլաստերավորման ժամանակ կլաստերները արտացոլվում են հիերարխիկ կառուցվածքով, դենդրոգրամի տեսքով(այլ կերպ՝ ծառի տեսքով): Դենդրոգրամի մեջ յուրաքանչյուր ճյուղ ներկայացնում է իրենից մի կլաստեր: Հիերարխիկ կլաստերավորման օրինակ է՝ Բաշխման վրա հիմնված կլաստերավորումը(Distribution based clustering):
- Կենտրոնի գաղափարի վրա հիմնված մեթոդ(Centroid-based clustering algorithms)-այն տվյալների բազան բաժանում է նմանատիպ տվյալներ պարունակող խմբերի, որոնք հայտնի են որպես կլաստերներ: Այս տեխնիկան օգտագործում է կենտրոններ, այնպես որ տվյալների կետերը հնարավորինս մոտ լինեն կլաստերի կենտրոնին: Կենտրոնի գաղափարը օգտագործող կլաստերավորման ամենահայտնի ալգորիթմը K-means կլաստերավորումն է:
- Խտության վրա հիմնված կլաստերավորում(Density-based clustering)-այն միավորում է մեծ խտություն ունեցող հատվածները կլաստերների մեջ: Խտության վրա հիմնված ամենատարածված մեթոդը՝ DBSCAN-(Density-Based Spatial Clustering of Applications with Noise)-ն է:



Հետազոտության շրջանակներում անդրադառնալու ենք «K-means» և «Գաուսյան Խառնուրդի Մոդել» մեթոդներին:

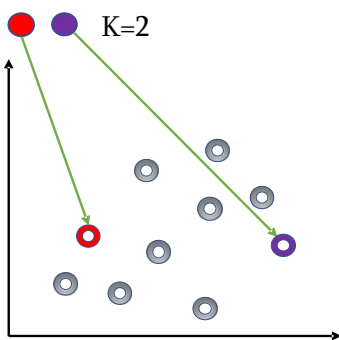
## 2.2 K-means ալգորիթմ

K-means կլաստերավորումը չվերահսկվող ուսուցման ալգորիթմ է, որն օգտագործվում է տվյալների կետերը բաժանելու առանձին խմբերի կամ կլաստերների՝ հիմնվելով դրանց նմանությունների վրա, այն լայնորեն օգտագործվում է տարբեր ոլորտներում, ներառյալ ֆինանսներում:

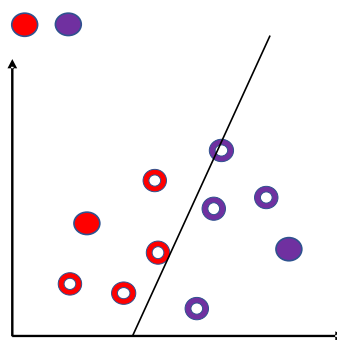
Որոշելով կլաստերները ֆինանսական տվյալների մեջ՝ k-means-ը կարող է օգնել շուկայական ռեժիմի նույնականացմանը, պորտֆելի օպտիմալացմանը, հաճախորդների սեգմենտավորմանը և այլ առաջադրանքներին, որոնք պահանջում են տվյալների կետերի խմբավորում՝ հիմնվելով դրանց բնութագրերի վրա:

K-means-ը կենտրոնի գաղափարը օգտագործող մեթոդ է, որը անընդմեջ ճշգրտում է կետերի կլաստերների համարները՝ նվազագույնի հասցնելով յուրաքանչյուր տվյալների կետի և այն կլաստերի կենտրոնի միջև եղած քառակուսի հեռավորությունների գումարը, որին պատկանում է:

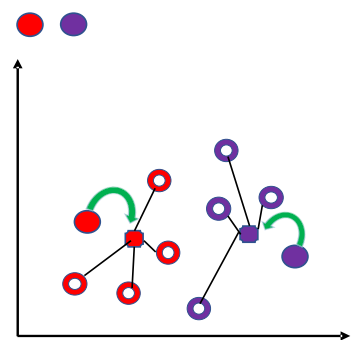
Գծապատկեր 3: K-means



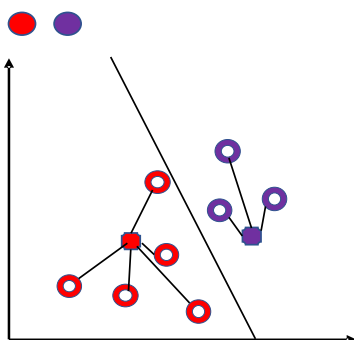
Գծապատկեր 4: K-means



Գծապատկեր 5: K-means



Գծապատկեր 6: K-means



Ներկայացնենք ավելի պատկերավոր: Ի սկզբանե մենք ֆիքսում ենք կլաստերների քանակը, ենթադրենք  $K=2$ , այնուհետև համակարգը տվյալների հավաքածուից պատահական կետեր է ընտրում, որպես կլաստերների կենտրոններ (Գծապատկեր 3):

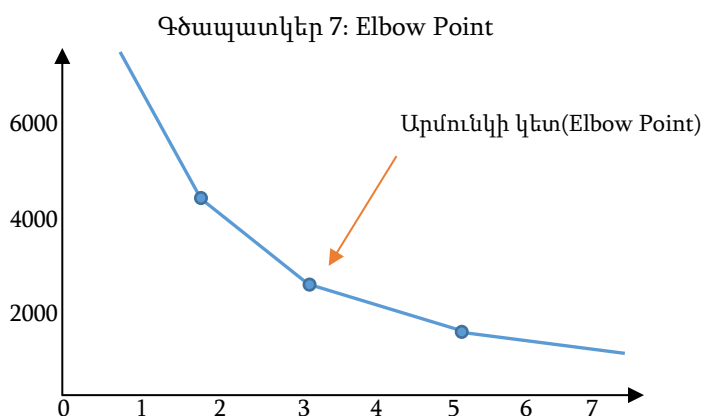
Հաջորդ քայլում տեսնում ենք թե ինչ հեռավորության վրա են գտնվում տվյալների կետերը, այսպես ասած ընտրված կենտրոնների նկատմամբ, հաշվարկելով Էվկլիդեսյան հեռավորությունները (Գծապատկեր 4), այնուհետև ալգորիթմը խմբավորում է այդ կետերը (Գծապատկերում նշված է միևնույն գույներով): Որից հետո արդեն իսկ խմբավորված կետերի համար ընտրում է կետրոն (centroid) և հեռավորությունը տվյալների կետերի հաշվարկում է այդ կենտրոնի նկատմամբ (Գծապատկեր 5): Գործողությունը կատարվում է այնքան ժամանակ մինչև կենտրոնները կայունանում են, այսինքն դրանց արժեքներում փոփոխություն չի նկատվում, քանի որ կլաստերը հաջող է ստացվել (Գծապատկեր 6):

Հիմնական խնդիրը կլաստերների ( $K$ ) քանակի ընտրությունն է: Տարբեր մեթոդներ կան կլաստերների օպտիմալ թվի ընտրության համար:

Արմունկի մեթոդը (Elbow method) առաջարկում է տարբեր կլաստերների քանակ ֆիքսել և այդ կլաստերների համար հաշվարկել տվյալ մեծությունը՝

$$J = \sum_{j=1}^K \sum_{i=1}^n \left\| x_i^{(j)} - c_j \right\|^2 \quad (1)$$

Որտեղ  $c_j$ -ն  $j$ -րդ կլաստերի կենտրոնն է (միջինը),  $x_i^{(j)}$ -երը  $j$ -րդ կլաստերին պատկանող կետերն են: (1) արտահայտությունը ցույց է տալիս թե որքան խիտ են մեր կլաստերները, այլ կերպ ասած, որքան մոտիկ են գտնվում կլաստերների մեջի կետերը այդ կլաստերների կենտրոններին: Մենք փորձում ենք հասկանալ թե որ  $K$ -ի դեպքում ենք ունենում խիտ կլաստերներ:



Եթե կլաստերների քանակը շատացնում ենք  $J$ -մեծությունը նվազում է (Գծապատկեր 7): Ընտրում ենք այն կետը որտեղից կտրուկ փոփոխություն ենք նկատում:

K-means-ի առավելությունն այն է, որ ալգորիթմի գործընթացը արագ է տեղի ունենում, հեշտ է մեկնաբանել ստացված արդյունքները, իսկ բացասական կողմերից է այն, որ պետք է ի սկզբանե ֆիքսել K-ի արժեքը, K-means-ը նաև զգայուն է outlier-ների նկատմամբ:

K-means-ը բավականին տարածված ալգորիթմ է, այն կարելի է կիրառել հաճախորդների սեգմենտավորման, պատկերների սեգմենտավորման, անոմալիաների հայտնաբերման և բազմաթիվ տարբեր տեսակի տվյալների և խնդիրների համար:

## 2.3 Գաուսյան Խառնուրդի Մոդել

Գաուսյան Խառնուրդի Մոդելը (Gaussian Mixture Model-GMM) մեթոդ է, որն օգտագործվում է հավանականակային կլաստերավորման մեջ: Մոդելները ենթադրում են, որ կա Գաուսի բաշխումների ֆիքսված քանակ, որոնցից յուրաքանչյուրը ներկայացնում է առանձին կլաստեր:

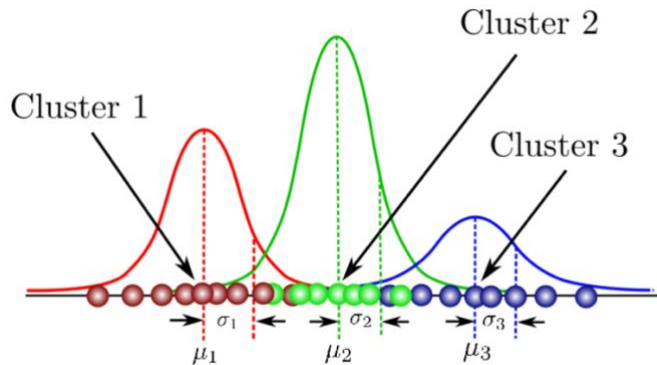
Ենթադրենք, որ ունենք մարդկանց հասակների տվյալների հավաքածու: Gaussian Mixture Model-ը կարող է օգտագործվել մարդկանց տարբեր ենթախմբերի բաժանելու համար՝ ելնելով նրանց հասակից: Բաշխումների վրա հիմնված տվյալները խմբավորելով՝ մոդելը կարող է բացահայտել մարդկանց ենթախմբերը, ովքեր համեմատաբար բարձրահասակ են, համեմատաբար ցածրահասակ կամ միջին հասակի:

Յուրաքանչյուր Գաուսի Խառնուրդի Մոդելը բաղկացած է հետևյալ պարամետրերից՝

- Կլաստերների քանակը, նշվում է «K»-ով, սա Գաուսյան բաշխումների թիվն է, որն օգտագործվում է տվյալների մոդելավորման համար: Պետք է նախապես նշել այս արժեքը,
- $\mu$  – միջինը, որը սահմանում է նրա կենտրոնը,

- կովարիացիոն մատրից - նշվում է  $\Sigma$ -ով, սա  $K$ -չափանի մատրից է, որը սահմանում է Գաուսի յուրաքանչյուր բաշխման ձևը:
- Խառնուրդի  $\pi$  հավանականություն - յուրաքանչյուր գաուսյանից լինելու հավանականությունը:

Գծապատկեր 8.6 Գաուսյան Խառնուրդի Մոդել



Գծապատկեր 8-ը նկարագրում է Գաուսյան Խառնուրդի Մոդելը:

GMM-ի առավելությունը  $K$ -Means-ի նկատմամբ այն է, որ  $K$ -Means-ը հաշվի չի առնում վարիացիան (այն վերաբերում է զանգի կորի ձևի լայնությանը), իսկ GMM-ն վերադարձնում է հավանականությունը, որ տվյալների կետերը պատկանում են  $K$  կլաստերներից յուրաքանչյուրին: GMM-ն օգտագործում է հավանականական մոտեցում և ապահովում է հավանականություն յուրաքանչյուր տվյալների կետի համար, որը պատկանում է կլաստերներին:

GMM-ի թերություններից մեկը՝ մոդելի ընտրությունն է, եթե պարամետրների թիվը չափազանց փոքր է, մոդելը կարող է ցույց չտալ տվյալների հիմքում ընկած բոլոր օրինաչափությունները:

GMM-ն լայնորեն կիրառվում է կլաստերավորման, անոմալիաների հայտնաբերման, ֆինանսներում և շատ ու շատ ոլորտներում: Ընդհանուր առմամբ, GMM-ները բազմակողմանի և հզոր մոդելավորման գործիք են, որոնք կարող են կիրառվել տարբեր տեսակի տվյալների և խնդիրների համար:

### ԳԼՈՒԽ 3. ԹԱՔՆՎԱԾ ՄԱՐԿՈՎՑԱՆ ՄՈԴԵԼՆԵՐ

#### 3.1 Մարկովյան շղթա

Մինչ Թաքնված Մարկովյան Մոդելները քննարկելը, անհրաժեշտ է պատկերացում կազմել Մարկովյան շղթայի մասին:

Մարկովյան շղթան<sup>7</sup> մոդել է, որը մեզ ինչ-որ բան է պատմում պատահական մեծությունների, վիճակների հաջորդականությունների հավանականությունների մասին, որոնցից յուրաքանչյուրը կարող է արժեքներ ընդունել որոշակի բազմությունից: Այս բազմությունները կարող են լինել բառեր կամ պիտակներ, որոնք ներկայացնում են որևէ բան, օրինակ՝ եղանակը: Շատ դեպքերում, մեզ հետաքրքրող իրադարձությունները թաքնված են և ուղղակիորեն չեն դիտարկվում:

Մարկովյան շղթայի հիմնական ենթադրությունը այն է, որ յուրաքանչյուր հաջորդ վիճակը գուշակելու համար, բավական է դիտարկել միայն ներկա վիճակը:

Մաթեմատիկորեն ներկայացնելու համար, ենթադրենք կան հետևյալ վիճակի (state) փոփոխականները՝  $q_1, q_2, \dots, q_i$ , ապա Մարկովյան հավասարումը կլինի՝

$$P(q_i = a | q_1 \dots q_{i-1}) = P(q_i = a | q_{i-1})$$

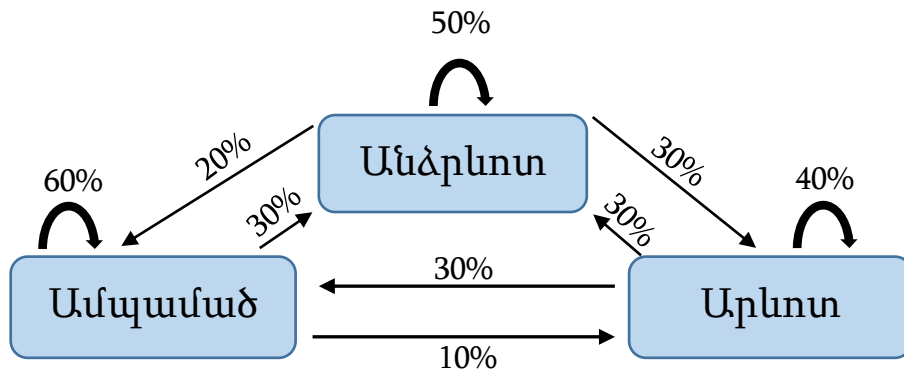
Այլ կերպ ասած, հավասարումը նշանակում է, որ ներկա վիճակի հավանականությունը՝ բոլոր նախորդ վիճակների պայմանով, նույնն է, ինչ ներկա վիճակի հավանականությունը, միայն նախորդ վիճակի պայմանով:

Մարկովյան շղթան նկարագրվում է հետևյալ բաղադրիչներով՝

- $Q = q_1, q_2, \dots, q_N$  -  $N$  վիճակների բազմություն,
- $A = a_{11}, a_{12}, \dots, a_{n1}, \dots, a_{nn}$  - անցումային հավանականության մատրից՝  $A$ , յուրաքանչյուր  $a_{ij}$ , ներկայացնում է  $i$  վիճակից  $j$  վիճակ տեղափոխվելու հավանականությունը, այնպես որ՝  $\sum_{j=1}^n a_{ij} = 1 \quad \forall i$ ,
- $\pi = \pi_1, \pi_2, \dots, \pi_N$  - վիճակների սկզբնական հավանականության բաշխում,  $\pi_i$ -ն հավանականությունն է, որ Մարկովյան շղթան կսկսի  $i$ -րդ վիճակից: Որոշ  $j$ -վիճակների համար կարող է  $\pi_j = 0$  լինել, սա նշանակում է, որ նրանք չեն կարող լինել սկզբնական վիճակներ, իսկ  $\sum_{i=1}^N \pi_i = 1$  :

Օրինակ<sup>8</sup>, դիտարկենք հնարավոր եղանակային պայմանները՝ արևոտ, անձրևոտ, ամպամած, այնուհետև կարող ենք ստեղծել Մարկովյան շղթա՝ երեք վիճակներով, որոնք համապատասխանում են այս եղանակային պայմաններին:

Գծապատկեր 9: Մարկովյան շղթայի օրինակ



Ըստ Գծապատկեր 9-ի անցումային հավանականությունները հետևյալն են՝

- Եթե այսօր արև է, ապա վաղը՝
  - 40% հավանականությամբ արև կլինի
  - 30% հավանականությամբ անձրև կլինի
  - 30% հավանականությամբ ամպամած կլինի
- Եթե այսօր անձրև է, ապա վաղը՝
  - 30% հավանականությամբ արև կլինի
  - 50% հավանականությամբ անձրև կլինի
  - 20% հավանականությամբ ամպամած կլինի
- Եթե այսօր ամպամած է, ապա վաղը՝
  - 10% հավանականությամբ արև կլինի
  - 30% հավանականությամբ անձրև կլինի
  - 60% հավանականությամբ ամպամած կլինի

Օգտագործելով այս Մարկովյան շղթան, կարող ենք հաշվարկել հավանականությունը, որ չորեքշաբթի կլինի ամպամած, եթե այսօր երկուշաբթի է և արևոտ:

1. Արևոտ – Արևոտ (երեքշաբթի) – Ամպամած (չորեքշաբթի). Ամպամած չորեքշաբթի օրվա հավանականությունը կարելի է հաշվարկել որպես  $0,4 \times 0,3 = 0,12$
2. Արևոտ – անձրևոտ (երեքշաբթի) – ամպամած (չորեքշաբթի). Ամպամած չորեքշաբթի օրվա հավանականությունը կարելի է հաշվարկել  $0,3 \times 0,2 = 0,06$ :

3. Արևոտ – Ամպամած (երեքշաբթի) – Ամպամած (չորեքշաբթի). Ամպամած չորեքշաբթի օրվա հավանականությունը կարելի է հաշվարկել  $0,3 \times 0,6 = 0,18$ :  
 Ամպամած չորեքշաբթի օրվա ընդհանուր հավանականությունը կլինի՝  $0,12 + 0,06 + 0,18 = 0,36$ :

### 3.2 Թաքնված Մարկովյան Մոդելներ

Շուկայական ռեժիմների հայտնաբերման հիմնական մեթոդը վիճակագրական ժամանակային շարքերի տեխնիկայի օգտագործումն է, որը հայտնի է որպես Թաքնված Մարկովյան Մոդել (անգլերեն՝ Hidden Markov Model(HMM)): HMM-ը վիճակագրական մոդել է, որում մոդելավորվող համակարգը Մարկովյան գործընթացներն են՝ չդիտարկվող կամ թաքնված վիճակներով: Մենք պիտակները անվանում ենք թաքնված, քանի որ դրանք չեն դիտարկվում:

HMM-ն նշվում է հետևյալ բաղադրիչներով՝

- $Q = q_1, q_2, \dots, q_N$  -  $N$  վիճակների բազմություն.
- $A = a_{11}, a_{12}, \dots, a_{n1}, \dots, a_{nn}$  - անցումային հավանականության մատրից՝  $A$ , յուրաքանչյուր  $a_{ij}$ , ներկայացնում է  $i$  վիճակից  $j$  վիճակ տեղափոխվելու հավանականությունը, այնպես որ՝  $\sum_{j=1}^n a_{ij} = 1 \quad \forall i$ .
- $O = o_1, o_2, \dots, o_T$  -  $T$  դիտարկումների հաջորդականություն, որոնցից յուրաքանչյուրը վերցված է  $V = v_1, v_2, \dots, v_V$  բառարանից.
- $B = b_i(o_t)$  -  $i$  վիճակից  $o_t$  դիտարկման առաջացման հավանականությունը,
- $\pi = \pi_1, \pi_2, \dots, \pi_N$  - վիճակների սկզբնական հավանականության բաշխում,  $\pi_i$ -ն հավանականությունն է, որ Մարկովյան շղթան կսկսի  $i$ -րդ վիճակից: Որոշ  $j$ -վիճակների համար կարող է  $\pi_j = 0$  լինել, սա նշանակում է, որ նրանք չեն կարող լինել սկզբնական վիճակներ, իսկ  $\sum_{i=1}^N \pi_i = 1$ .

Թաքնված Մարկովյան Մոդելը ունի երկու պարզ ենթադրություն՝

1. Ինչպես Մարկովյան շղթայում, այնպես և այստեղ, որոշակի վիճակի հավանականությունը կախված է միայն նախորդ վիճակից՝

$$P(q_i = a | q_1 \dots q_{i-1}) = P(q_i = a | q_{i-1})$$

2.  $o_i$  էլքի դիտարկման հավանականությունը կախված է միայն  $i$ -րդ վիճակից՝

$$P(o_i | q_1 \dots q_i, \dots, q_T, o_1, \dots, o_i, \dots, o_T) = P(o_i | q_i)$$

Թաքնված Մարկովյան Մոդելները ուսումնասիրում են երեք հիմնական խնդիրներ՝

- Ճշմարտանմանություն (Likelihood) – հաշվի առնելով Թաքնված Մարկովյան Մոդելների հետևյալ պարամետրները՝  $\lambda = (A, B)$  և դիտարկումների  $O$ -հաջորդականությունը, պետք է որոշել հավանականությունը՝  $P(O|\lambda)$ : Այն որոշվում է տարբեր ալգորիթմների միջոցով, ինչպիսիք են՝ Forward, Backward ալգորիթմները:
- Վերծանում (Decoding) – Սա պատասխանում է այն հարցին, թե  $n^{\circ}$ րն է ամենահավանական թաքնված վիճակների հաջորդականությունը, երբ ունենք դիտարկումների հաջորդականություն: Այսինքն հաշվի առնելով դիտարկումների  $O$ -հաջորդականությունը և HMM-ի՝  $\lambda = (A, B)$ , հայտնաբերել լավագույն թաքնված վիճակի հաջորդականությունը՝  $Q$ -ն (Viterbi ալգորիթմ):
- Ուսուցում (Learning) – Ի՞նչ կարող ենք սովորել մեր ունեցած տվյալներից: Այսինքն՝ դիտարկված տվյալներից ինչպե՞ս ստեղծել Թաքնված Մարկովյան Մոդելներ: Հաշվի առնելով դիտարկումների  $O$ -հաջորդականությունը և HMM-ի վիճակների բազմությունը, սովորենք HMM-ի՝  $A$  և  $B$  պարամետրերը (Baum-Welch ալգորիթմ):

Հետազոտության համատեքստում թաքնված վիճակները՝ շուկայի ռեժիմներն են, իսկ դիտարկելի տվյալները S&P 500-ի և VIX-ի ժամանակային շարքերն են: Մենք փորձելու ենք բացահայտել այդ ռեժիմները Python-ծրագրավորման լեզվի միջոցով, որտեղ ներդրված է հատուկ գրադարան, Թաքնված Մարկովյան Մոդելները օգտագործելու համար: Ըստ էության ալգորիթմը սկզբում սովորում է  $A$  և  $B$  պարամետրերը (Learning), այնուհետև հայտնաբերում է թաքնված վիճակների հաջորդականությունը (Decoding):



### 3.2.1 Թաքնված Մարկովյան Մոդելի Ուսուցում

Թաքնված Մարկովյան Մոդելը, հաշվի առնելով դիտարկումների Օ-հաջորդականությունը և HMM-ի վիճակների բազմությունը, սովորում է՝  $A$  և  $B$  պարամետրերը: Ուսուցման մեջ մենք ունենք չպիտակավորված դիտարկումների Օ-հաջորդականություն և թաքնված վիճակներ՝  $Q$ : Մենք չգիտենք թաքնված վիճակներից որևէ մեկում գտնվելու հավանականությունը: Թաքնված Մարկովյան Մոդելում ուսուցումը սկսում է մոդելի պարամետրները գուշակելուց, դրանք են՝ անցումային և դիտարկումների հավանականությունները: Այնուհետև այդ պարամետրների գնահատականները կրկնակի ճշգրտվում են:

Սահմանենք Forward՝ հավանականությունը՝

$$a_t(j) = P(o_1, o_2 \dots o_t, q_t = j | \lambda)$$

Հավասարումը ցույց է տալիս  $j$  վիճակում գտնվելու հավանականությունը առաջին  $t$  դիտարկումները տեսնելուց հետո:  $q_t = j$  – ն նշանակում է վիճակների հաջորդականության  $t$ -րդ վիճակը  $j$  վիճակն է: Այս հավանականությունը հաշվարկվում է հետևյալ կերպ՝

$$a_t(j) = \sum_{i=1}^N a_{t-1}(i) a_{ij} b_j(o_t),$$

Որտեղ որ  $a_{t-1}(i)$  –ն սահմանվում է որպես  $t-1$  պահին  $i$ -րդ վիճակում հավանականությունը,  $a_{ij}$ -ն ներկայացնում է  $i$  վիճակից  $j$  վիճակ տեղափոխվելու հավանականությունը,  $b_j(o_t)$  – ն  $i$  վիճակից  $o_t$  դիտարկման առաջացման հավանականությունը:

Ալգորիթմը հասկանալու համար սահմանենք backward հավանականությունը(backward probability)՝

$$\beta_t(i) = P(o_{t+1}, o_{t+2} \dots o_T | q_t = i, \lambda),$$

որտեղ  $\beta$ -ն  $t+1$ -պահից մինչև վերջ դիտարկումները տեսնելու հավանականությունն է հաշվի առնելով, որ  $t$  պահին գտնվում ենք  $i$ -րդ վիճակում:

Հաշվարկման քայլերը հետևյալն են՝

- Սկզբնավորում(Initialization)՝

$$\beta_T(i) = 1, \quad 1 \leq i \leq N$$

Այն ցույց է տալիս, որ մենք գտնվում ենք վերջնական վիճակում և հավասար է 1-ի, քանի որ մենք դիտարկել ենք ամբողջ հաջորդականությունը մինչև T-պահիր:

- Ռեկուրսիա(Recursion)`

$$\beta_t(i) = \sum_{j=1}^N a_{ij} b_j(o_{t+1}) \beta_{t+1}(j), \quad 1 \leq i \leq N, 1 \leq t < T$$

Այն ներկայացնում է t+1-ից մինչև T պահիր դիտարկումների հաջորդականությունը դիտարկելու հավանականությունը, հաշվի առնելով, որ HMM-ը գտնվում է t-պահին i-րդ վիճակում:

- Ավարտ(Termination)`

$$P(O|\lambda) = \sum_{j=1}^N \pi_j b_j(o_1) \beta_1(j)$$

Սա ցույց է տալիս O-դիտարկումների հաջորդականության հավանականությունը:

Տեսնենք թե ինչպես կարող ենք օգտագործել forward և backward հավանականությունները դիտարկման հաջորդականությունից հաշվարկելու անցման հավանականությունները, չնայած որ մոդելի միջոցով անցած իրական ճանապարհը թաքնված է: Թաքնված Մարկովյան մոդելներում վիճակների միջև անցման հավանականությունը գնահատելու համար մենք կարող ենք օգտագործել պարզագույն առավելագույն ճշմարտանմանության(Likelihood) գնահատման տարբերակը: Կարող ենք գնահատել i վիճակից j վիճակի անցնելու հավանականությունը հետևյալ կերպ`

$$\hat{a}_{ij} = \frac{i \text{ վիճակից } j \text{ վիճակ անցումների ակնկալվող թիվը}}{i \text{ վիճակից անցումների ակնկալվող թիվը}}$$

i վիճակից j վիճակի անցումների ակնկալվող թիվը հաշվարկելու համար մենք պետք է հաշվի առնենք այն հավանականությունը, որ այս անցումը կատարվել է դիտարկման հաջորդականության t ժամանակի յուրաքանչյուր կետում: Գումարելով t բոլոր ժամանակահատվածները, կարող ենք գնահատել  $i \rightarrow j$  անցման ընդհանուր թիվը:

Օրինակ, ենթադրենք, որ մենք ունենք դիտարկումների հաջորդականություն, որը բաղկացած է բաժնետոմսերի օրական գներից, և մենք ցանկանում ենք գնահատել շուկայական տարբեր ռեժիմների միջև անցման հավանականությունը: Մենք կարող ենք սահմանել երկու վիճակ` «ցուլ» և «արջ», և օգտագործել Թաքնված Մարկովյան Մոդելը` այս վիճակների միջև անցումները մոդելավորելու համար` հիմնվելով բաժնետոմսերի օրական գների վրա: Անցման հավանականությունները գնահատելու

համար մենք կարող ենք հաշվարկել  $i$  վիճակից  $j$  վիճակի անցումների ակնկալվող թիվը՝ գումարելով  $t$  բոլոր ժամանակային կետերը, որտեղ մոդելը կանխատեսում է  $i$  վիճակից  $j$  վիճակի անցում: Այս ակնկալվող թիվը պետք է բաժանել  $i$  վիճակից անցումների ընդհանուր ակնկալվող թվի վրա՝ ստանալու համար  $\hat{a}_{ij}$  անցման հավանականությունը:

Սահմանենք հետևյալ փոփոխականը՝  $\xi_t$  – այն ցույց է տալիս, որ  $t$  պահին գտնվում ենք  $i$  վիճակում, իսկ  $t + 1$  պահին  $j$  վիճակում՝

$$\xi_t(i, j) = P(q_t = i, q_{t+1} = j | O, \lambda), \quad (2)$$

(2)-րդ հավասարումը հաշվարկելու համար, մենք սկզբում հաշվում ենք հետևյալ արտահայտությունը՝

$$\xi_t(i, j) = P(q_t = i, q_{t+1} = j, O | \lambda), \quad (3)$$

(3)-ը ցույց է տալիս  $t$  ժամանակում  $i$  վիճակում գտնվելու,  $t+1$  պահին  $j$  վիճակին անցնելու և ամբողջ հաջորդականությունը դիտարկելու համատեղ հավանականությունը: (3) արտահայտության հաշվարկը հետևյալն է՝

$$\xi_t(i, j) = a_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j),$$

(2)-րդ բանաձևը հաշվարկելու համար հետևում ենք հավանականության օրենքներին, այն է որ՝

$$P(X|Y, Z) = \frac{P(X, Y|Z)}{P(Y|Z)}$$

Իսկ  $P(O|\lambda)$  ներկայացնում ենք այսպես՝

$$\sum_{j=1}^N a_t(j) \beta_t(j)$$

Այսպիսով վերջնական  $\xi_t$  – ն կլինի հետևյալ կերպ՝

$$\xi_t(i, j) = \frac{a_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)}{\sum_{j=1}^N a_t(j) \beta_t(j)}$$

Այս դեպքում  $i$ -վիճակից  $j$ -վիճակի անցման ակնկալվող թիվը կլինի բոլոր  $t$ -ժամանակային քայլերը գումարելով:  $\hat{a}_{ij}$ -ն հաշվարկելու համար նաև պետք է հաշվել  $i$  վիճակից անցումների ընդհանուր ակնկալվող թիվը, ուստի գումարում ենք բոլոր անցումները  $i$ -րդ վիճակից՝

$$\hat{a}_{ij} = \frac{\sum_{t=1}^{T-1} \xi_t(i, j)}{\sum_{t=1}^{T-1} \sum_{k=1}^N \xi_t(i, k)} \quad (4)$$

Մեզ անհրաժեշտ է նաև դիտարկումների հավանականությունը վերահաշվարկելու բանաձև: Դիտարկումների հավանականությունը ներկայացնում է որոշակի սիմվոլ՝  $v_k$  (վերցված  $V$  բառարանից) դիտարկելու հավանականությունը, հաշվի առնելով հիմքում ընկած վիճակը: Վերահաշվարկի բանաձևը կլինի հետևյալը՝

$$\hat{b}_j(v_k) = \frac{\text{ակնկալվող } j \text{ վիճակի քանակը և } v_k \text{ սիմվոլը}}{\text{ակնկալվող } j \text{ վիճակի քանակը}}$$

Մեզ անհրաժեշտ է իմանալ  $t$  պահին  $j$  վիճակում գտնվելու հավանականությունը, որը մենք կանվանենք՝  $\gamma_t(j)$ , բանաձևը հետևյալն է՝

$$\gamma_t(j) = P(q_t = j | O, \lambda)$$

Մեկ անգամ ևս կհաշվենք՝ ներառելով դիտարկման հաջորդականությունը հավանականության մեջ՝

$$\gamma_t(j) = \frac{P(q_t = j, O | \lambda)}{P(O | \lambda)} \quad (5)$$

(5)-ը ներկայացնենք հետևյալ կերպ՝

$$\gamma_t(j) = \frac{a_t(j)\beta_t(j)}{P(O | \lambda)}$$

Այնուհետև կարող ենք հաշվարկել  $b$ -ն(դիտարկման հավանականությունը)՝

$$\hat{b}_j(v_k) = \frac{\sum_{t=1}^T s.t. O_t=v_k \gamma_t(j)}{\sum_{t=1}^T \gamma_t(j)} \quad (6)$$

(6) արտահայտության համարիչում մենք գումարում ենք  $\gamma_t(j)$ -ն բոլոր ժամանակային քայլերի համար, որտեղ որ  $O_t$  դիտարկումը մեզ հետաքրքրող  $v_k$  նշանն է: Հայտարարում գումարում ենք  $\gamma_t(j)$ -ն բոլոր  $t$  ժամանակային քայլերի համար: Մենք արդեն ունենք (4) և (6) հավասարումները  $A$ -ի անցման և  $B$  դիտարկման հավանականությունները վերագնահատելու  $O$  դիտարկումների հաջորդականությունից՝ ենթադրելով, որ մենք արդեն ունենք  $A$ -ի և  $B$ -ի նախկին գնահատականները:

Ամփոփելով ասենք, որ ալգորիթմը բաղկացած է երկու քայլից, առաջին քայլում հաշվարկվում է դիտարկումների հաջորդականության յուրաքանչյուր փուլում յուրաքանչյուր վիճակում գտնվելու հավանականությունը՝  $\gamma$ , և դիտարկումների հաջորդականության մեկ վիճակից մյուսին անցնելու հավանականությունը՝  $\xi$ , իսկ երկրորդ քայլում օգտագործում ենք  $\gamma$ -ն և  $\xi$ -ը վերագնահատելու անցումը և դիտարկումների հաջորդականությունը:

### 3.2.2 Թաքնված Մարկովյան Մոդելի Վերծանում

Վերծանման ժամանակ ունենք դիտարկումների Օ-հաջորդականություն և HMM-ի՝  $\lambda = (A, B)$ , պետք է հայտնաբերել լավագույն թաքնված վիճակի հաջորդականությունը՝  $Q$ -ն: Խնդրի լուծման համար օգտագործվում է Viterbi-ի ալգորիթմը, այն դինամիկ ծրագրավորման տեսակ է, իսկ դինամիկ ծրագրավորումը վերաբերում է բարդ խնդիրների պարզեցմանը՝ այն ավելի պարզ խնդիրների բաժանելով, ռեկուրսիվ եղանակով:

Viterbi-ի<sup>7</sup> ալգորիթմի ժամանակ հաշվարկվում է հետևյալ մեծությունը՝

$$v_t(j) = \max_{q_1, \dots, q_{t-1}} P(q_1 \dots q_{t-1}, o_1, o_2 \dots o_t, q_t = j | \lambda)$$

Վերը նշված բանաձևը ցույց է տալիս  $t$ -պահին  $j$ -վիճակում գտնվելու ամենամեծ հավանականությունը, հաշվի առնելով դիտարկումների հաջորդականությունը մինչև  $t$ -պահ: Viterbi-ն հաշվարկում է յուրաքանչյուր վիճակում գտնվելու հավանականությունը հաշվի առնելով, որ մենք արդեն հաշվարկել էինք  $t - 1$  պահին յուրաքանչյուր վիճակում գտնվելու հավանականությունը: Տրված  $q_j$  վիճակի համար  $t$  ժամանակում  $v_t(j)$  արժեքը հաշվարկվում է հետևյալ կերպ՝

$$v_t(j) = \max_{i=1 \dots N} v_{t-1}(i) a_{ij} b_j(o_t)$$

որտեղ՝

- $v_{t-1}(i)$  – ցույց է տալիս նախորդ ժամանակահատվածում  $i$ -վիճակում գտնվելու ամենամեծ հավանականությունը,
- $a_{ij}$  – ցույց է տալիս  $q_i$  վիճակից՝  $q_j$  անցնելու հավանականությունը,
- $b_j(o_t)$  –  $o_t$  դիտարկման հավանականությունը, հաշվի առնելով ներկայիս  $j$ -վիճակը: Երբ ալգորիթմը շարժվում է դեպի նոր վիճակ, ապա կարող ենք այդ վիճակից հետ քայլեր կատարել և հասնել այն լավագույն ճանապարհին, որը մեզ հասցրել էր դեպի այդ վիճակը:

Ներկայացնենք Viterbi-ի ալգորիթմի քայլերը հետևյալ կերպ՝

- Սկզբնավորում(Initialization)՝

$$v_1(j) = \pi_j b_j(o_1) \quad 1 \leq j \leq N$$

$$b_{t1}(j) = 0 \quad 1 \leq j \leq N$$

$t = 1$  պահին, մենք չունենք դիտարկվող նախկին վիճակներ, ուստի մենք բոլոր  $j$ -երի համար  $b_{t1}(j)$  -ին սահմանում ենք հավասար զրոյի: Մա ուղղակի պայման է և չի ազդում վիճակի ամենահավանական հաջորդականության հաշվարկի վրա: Մյուս կողմից,  $v_1(j)$  փոփոխականը ներկայացնում է  $j$  վիճակում գտնվելու հավանականությունը առաջին քայլում՝ հաշվի առնելով  $\pi$  վիճակի սկզբնական բաշխումը և առաջին դիտարկումը՝  $o_1$  :

• Ռեկուրսիա(Recursion)՝

$$v_t(j) = \max_{i=1..N} v_{t-1}(i) a_{ij} b_j(o_t); \quad 1 \leq j \leq N, 1 < t \leq T \quad (7)$$

$$bt_t(j) = \operatorname{argmax}_{i=1..N} v_{t-1}(i) a_{ij} b_j(o_t); \quad 1 \leq j \leq N, 1 < t \leq T \quad (8)$$

(7) արտահայտությունը հաշվարկում է  $t$ -պահին  $j$ -վիճակ տանող բոլոր հնարավոր ուղիների առավելագույն հավանականությունը:

(8) արտահայտությունը օգտագործվում է վիճակի ամենահավանական հաջորդականությունը վերականգնելու համար, երբ ավարտվի Viterbi-ի ալգորիթմը:

• Ավարտ(Termination)՝

$$\text{Լավագույն միավոր} \quad P^* = \max_{i=1..N} v_T(i) \quad (9)$$

$$\text{Հետադարձի սկիզբը} \quad QT^* = \operatorname{argmax}_{i=1..N} v_T(i) \quad (10)$$

(9) արտահայտությունը ներկայացնում է ամենահավանական թաքնված վիճակի հաջորդականության հավանականությունը՝ հաշվի առնելով դիտարկվող հաջորդականությունը:

(10) - ը ամենամեծ հավանականություն ունեցող վիճակն է, սա ամենահավանական հաջորդականության վերջին թաքնված վիճակն է:

Viterbi-ի ալգորիթմը լայնորեն օգտագործվում է խոսքի ճանաչման, բնական լեզվի մշակման, կենսահինֆորմատիկայի և շատ այլ ոլորտներում, որտեղ վերլուծվում են հաջորդական տվյալները: Այն արդյունավետ ալգորիթմ է HMM-ում թաքնված վիճակների ամենահավանական հաջորդականությունը գտնելու համար և համարվում է մեքենայական ուսուցման ոլորտում հիմնական ալգորիթմներից մեկը:

#### ԳԼՈՒԽ 4. ՎԵՐԼՈՒԾՈՒԹՅՈՒՆ

Այս գլխում մենք փորձում ենք հայտնաբերել շուկայի տարբեր ռեժիմներ՝ օգտագործելով S&P(Standard & Poor's) 500-ի փական գների, ծավալի և VIX(Volatility Index) ինդեքսի փական արժեքի 2018թ.-ից սկսած պատմական օրական տվյալները:

S&P 500-ը ֆոնդային շուկայի ինդեքս է, որը բաղկացած է Միացյալ Նահանգների 500 խոշոր ընկերություններից, այն լայնորեն համարվում է ԱՄՆ ֆոնդային շուկայի ընդհանուր կատարողականի չափանիշ: S&P 500 ինդեքսի փական գինը ինդեքսի վերջնական գինն է առևտրային օրվա վերջում, ինդեքսի փական գինը որոշվում է՝ հաշվի առնելով ինդեքսում ընդգրկված 500 ընկերությունների բաժնետոմսերի գների կշռված միջինը: S&P 500 ինդեքսի փական գինը կարող է օգտագործվել ԱՄՆ ֆոնդային շուկայի ընդհանուր կատարողականը վերլուծելու և շուկայական միտումների հիման վրա ներդրումային որոշումներ կայացնելու համար:

VIX-ը անկայունության չափանիշ է, ինդեքսի փական արժեքը ներկայացնում է շուկայի անկայունության ակնկալիքը առաջիկա 30 օրվա ընթացքում: VIX ինդեքսը լայնորեն կիրառվում է և համարվում է շուկայի անկայունության հուսալի ցուցիչ, այդ իսկ պատճառով այն հաճախ ներառվում է շուկայական ռեժիմների վերլուծություններում:

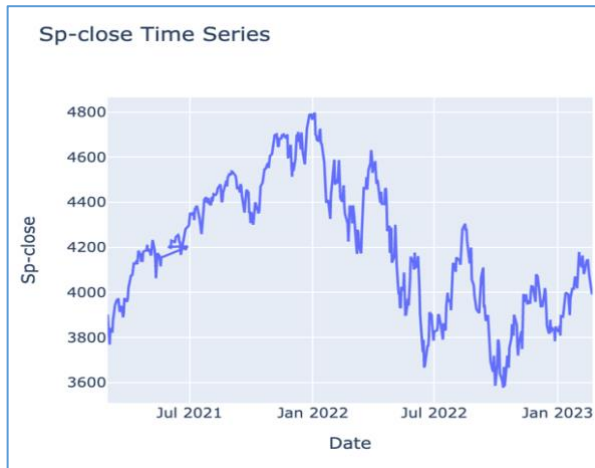
#### **Շուկայի ռեժիմների հայտնաբերում կլաստերավորման ալգորիթմների միջոցով.**

##### ***K-means ալգորիթմ***

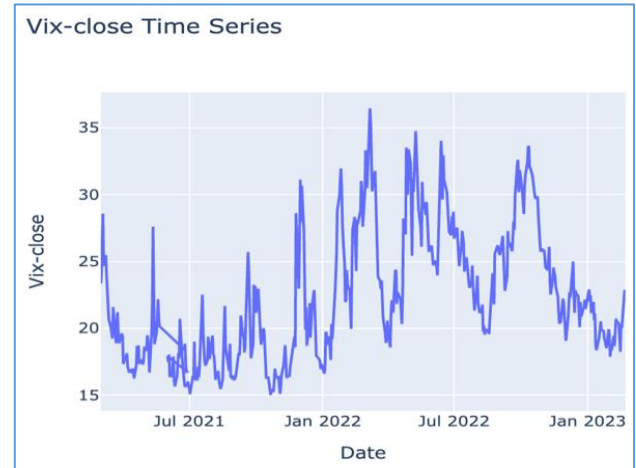
K-means ալգորիթմի կիրառման համար սկզբում օգտագործել ենք 2 տարվա տվյալների հավաքածու, որպեսզի հասկանանք նրա աշխատանքը տարբեր չափերի տվյալների հետ:

Կառուցենք S&P-ի և VIX-ի ժամանակային շարքը, հասկանալու համար շուկայի դինամիկան:

Գծապատկեր 10: S&P-close Time Series

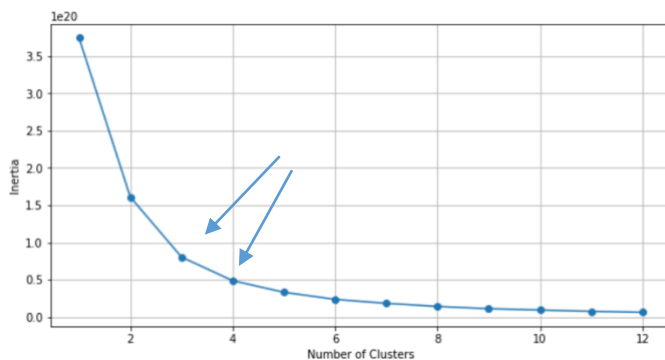


Գծապատկեր 11: VIX-close Time Series



Գծապատկեր 10-ը ցույց է տալիս, որ 2022-ից S&P-ի գները զգալիորեն ընկել են, իսկ Գծապատկեր 11-ում այդ նույն ժամանակահատվածում VIX-ը զգալիորեն բարձրացել է: Եվ իրոք որ 2022 թվականին S&P 500 ինդեքսը նվազման միտում է ունեցել մի քանի գործոնների պատճառով, այդ թվում՝ COVID-19 համաճարակի երկարատև հետևանքների, Ուկրաինա-Ռուսական հակամարտության հետևանքով էներգիայի խափանման և տնտեսական անկման մտահոգությունների պատճառով: Իսկ VIX ինդեքսի աճը վկայում է շուկայի անորոշության և ռիսկի ավելի բարձր մակարդակի մասին:

Գծապատկեր 12: Արմունկի կետ(Elbow point)



Գծապատկեր 12-ը ցույց է տալիս Elbow point-ը, որտեղ սլաքներով նշված կետերը օպտիմալ կլաստերների քանակներն են, քանի որ հենց այդտեղից է տեղի ունենում կտրուկ փոփոխություն, մենք կարող ենք դիտարկել 3 կամ 4 կլաստեր, ինչու չէ նաև 2, ի վերջո այնքան որքան ցանկանում ենք:

Կիրառելով ալգորիթմը ստանում ենք հետևյալ պատկերը՝



Գծապատկեր 13: Կլաստերավորելուց հետո



Գծապատկեր 13-ում տեսնում ենք, որ K-means-ը հաջողությամբ բաժանել է 3 կլաստերների՝ 0, 1 և 2: Նշենք, որ տվյալները նորմալիզացված են նախքան K-means-ի կիրառումը:

Այնուհետև հաշվարկում ենք միջինները որպեսզի հասկանանք թե որ շուկայական ռեժիմն է նկարագրում յուրաքանչյուր կլաստեր:

Աղյուսակ 3: Տվյալների միջինները

	Sp-close	Volume	Vix-close
kmeans_3			
0	4516.258258	4.125253e+09	19.294494
1	3971.844892	5.025191e+09	28.845252
2	4071.237389	4.240762e+09	20.337056

Տեսնում ենք որ ամենա ցածր տատանողականությունը և ամենա բարձր փակման գինը եղել է 0-ռեժիմում, իսկ ամենա բարձր տատանողականությունը և ամենա ցածր փակման գինը 1-ում, իսկ ռեժիմ 2-ը այդքան էլ չի տարբերվում 0-ից:

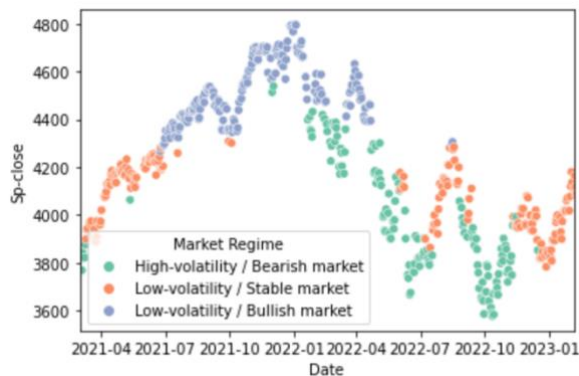
Այնուամենայնիվ սա բավարար չէ ռեժիմները պիտակավորելու համար: Այդ իսկ պատճառով մենք հաշվարկում ենք S&P-ի փակման գների եկամուտները(returns), որպեսզի տեսնենք թե որ ռեժիմն է ուղեկցվում ավելի շատ դրական եկամուտներով, որը ոչ:

Աղյուսակ 4: Returns

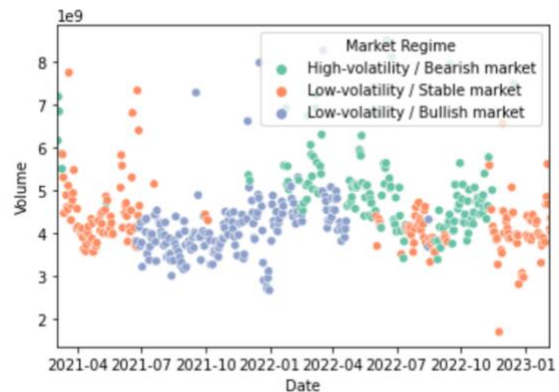
	mean	count	positive_returns	negative_returns	label
kmeans_3					
0	0.000937	178	104	74	Bullish
1	-0.002527	139	51	88	Bearish
2	0.001404	180	91	89	Bullish

Աղյուսակ 4-ում տեսնում ենք, որ 0-ռեժիմը ուղեկցվում է ավելի շատ դրական շահույթով, որին որ վերագրել ենք Bullish մարկետ, հաշվի առնելով նաև VIX-ի ցածր տատանողականությունը: Ռեժիմ 1-ը ուղեկցվում է ավելի շատ բացասական շահույթով, և այստեղ առկա է ամենա բարձր VIX -ի տատանողականությունը: Իսկ 2-րդ ռեժիմը մենք անվանել ենք Կայուն շուկա:

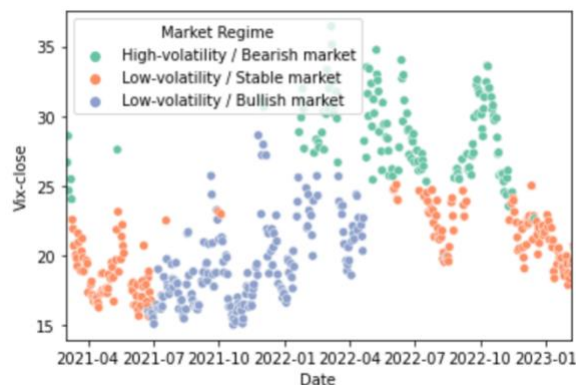
Գծապատկեր 14: S&P close price



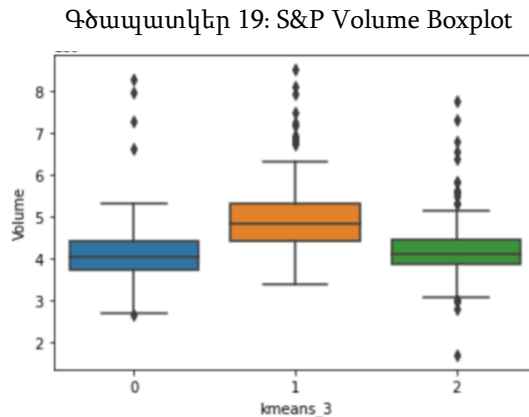
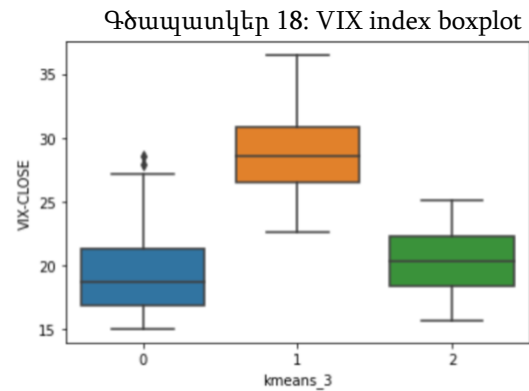
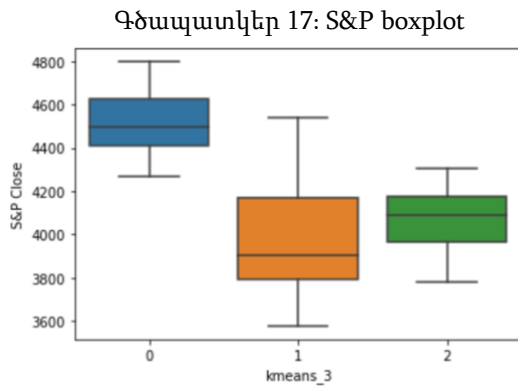
Գծապատկեր 15: S&P Volume



Գծապատկեր 16: VIX index



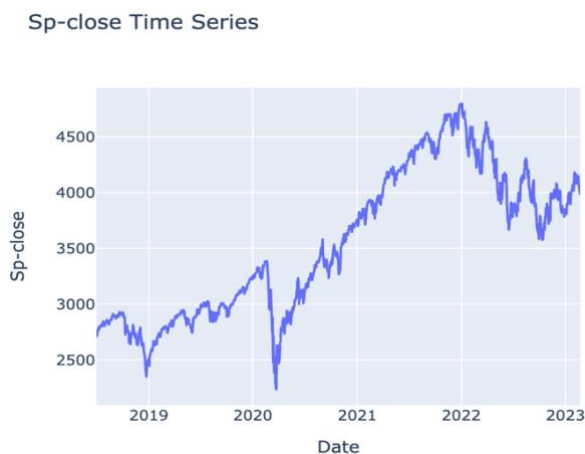
Այս գծապատկերները ներկայացնում են յուրաքանչյուր տվյալին համապատասխան շուկայական ռեժիմները: Իսկ Boxplot-երը հավաստիացնում են մեր արդյունքներին պիտակներ տալու ճշտությունը՝



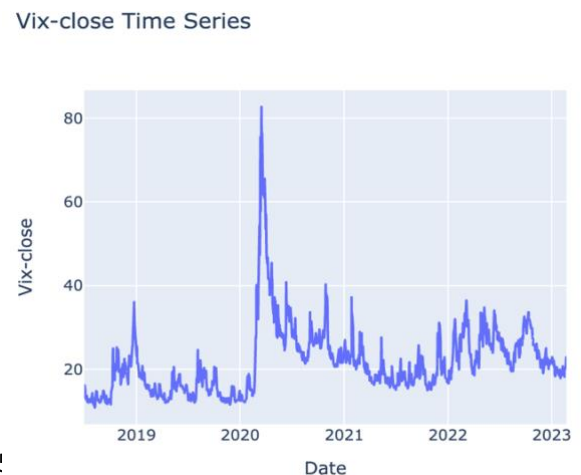
Boxplot-երից ակնհայտ երևում է, որ 1-ռեժիմում ամենա բարձր տատանողականությունն է եղել, որը ուղեկցվել է առևտրի ծավալի զգալի աճով, այն արտացոլում է դինամիկ և արագ փոփոխվող շուկայական միջավայրը, որտեղ ներդրողները ակտիվորեն արձագանքում են նոր տեղեկատվություններին և համապատասխանաբար ճշգրտում իրենց ներդրումային ռազմավարությունները:

Այնուհետև կիրառենք K-means-ը 5 տարվա տվյալների բազայի վրա, կրկին օգտագործելով նորմալիզացված տվյալներ:

Գծապատկեր 20: S&P-close Time Series

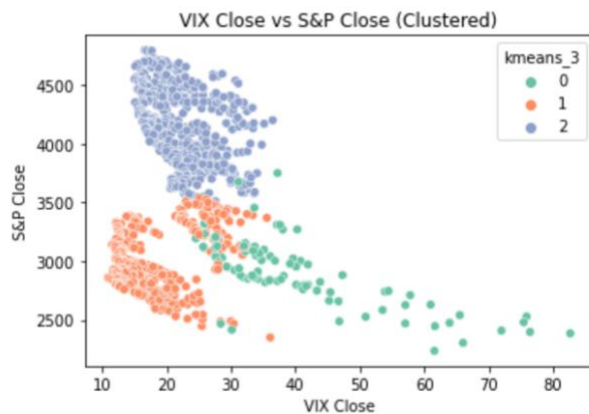


Գծապատկեր 21: VIX-close Times Series



Գծապատկեր 20-ից և 21-ից երևում է, որ 2018 թվականի դեկտեմբերին S&P 500-ի փակման գները անկում են ապրել, և ըստ պաշտոնական տվյալների S&P 500-ը ավարտել է 2018 թվականը ավելի քան 6% կորստով: 2020-ի կտրուկ անկումը COVID 19-ի հետևանք է, իսկ 2022-ինը ֆինանսական ճգնաժամի, այդ նույն ժամանակահատվածներում VIX-ինդեքսը կտրուկ տատանվել է, որն էլ ակնհայտ երևում է գծապատկերում:

Գծապատկեր 22: Clustering

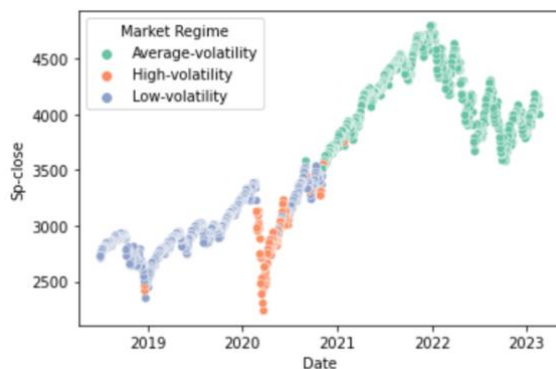


Գծապատկեր 22-ը բաժանել է երեք կլաստերների, տեսնենք թե ինչ արդյունքներ ունենք:

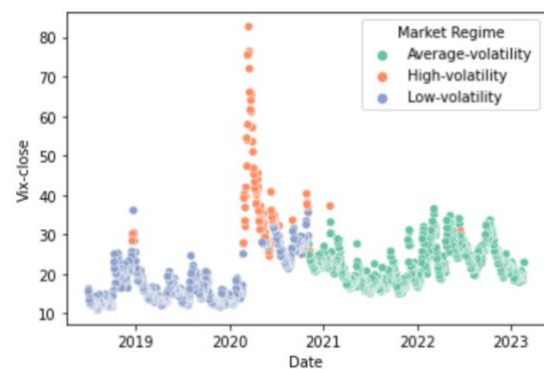
Աղյուսակ 5: Տվյալների միջինները

	Sp-close	Volume	Vix-close
kmeans_3			
0	2894.023295	6.436732e+09	40.763068
1	2981.044229	3.692418e+09	17.796759
2	4141.473049	4.483430e+09	22.451847

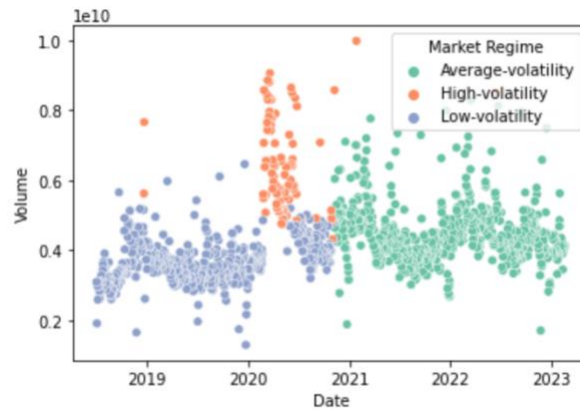
Գծապատկեր 23: S&P close price



Գծապատկեր 24: VIX index



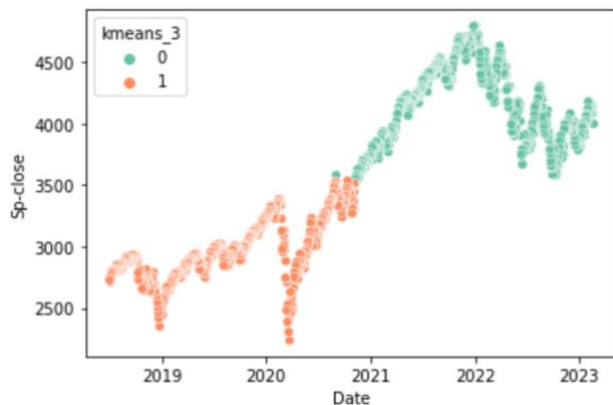
Գծապատկեր 25: S&P Volume



Այս գծապատկերներից տեսնում ենք, որ 2020-ի համաճարակի շրջանը այն դիտարկել է, որպես առանձին կլաստեր, որտեղ որ առկա է բարձր տատանողականություն և միևնույն ժամանակ առևտրի ծավալի զգալի աճ, կարող ենք ենթադրել որ մարդկանց մոտ այս ժամանակահատվածում եղել է խուճապ:

Այնուամենայնիվ K-means-ը կարողանում է ինչ-որ կերպ բաժանել 3 կլաստերների, սակայն 2 կլաստերի դեպքում այն անհաջող է կատարում բաժանումը՝

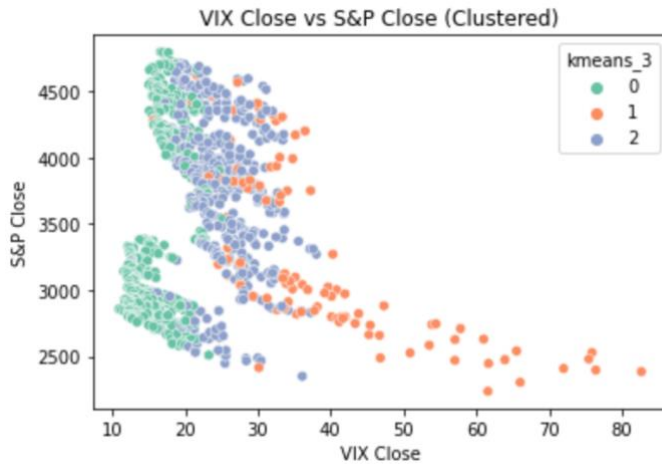
Գծապատկեր 26: S&P Close, with 2 clusters



Գծապատկեր 26-ից ակնհայտ երևում է, որ 2018-ից մինչև 2021 դիտարկել է առանձին ռեժիմ, իսկ 2021-ից սկսած մեկ այլ ռեժիմ, որը որ չի արտացոլում հստակ շուկայի դինամիկան: Երկու ռեժիմներն էլ այստեղ ունեն միջին տատանողականություն և միջին առևտրի ծավալ, ուստի մենք գրեթե ոչինչ չենք կարող եզրակացնել այս կլաստերներից:

Այնուհետև որպես նոր feature վերցրել ենք S&P 500-ի լոգարիթմական եկամտաբերությունը(log return), որպեսզի հասկանանք ալգորիթմը արձագանքը այս դեպքում:

Գծապատկեր 27: Clustering with S&P log return



Գծապատկեր 27-ում ալգորիթմը բաժանել է երեք կլաստերների, նկատենք որ S&P 500-ի փակման գների նորմալիզացված տարբերակը օգտագործելով, այն վիզուալ ավելի լավ էր բաժանում կլաստերների(Գծապատկեր (13), (22)): Ընդհանուր առմամբ K-means-ը կարող է ոչ այդքան լավ բաժանել կլաստերները հետևյալ պատճառներով՝

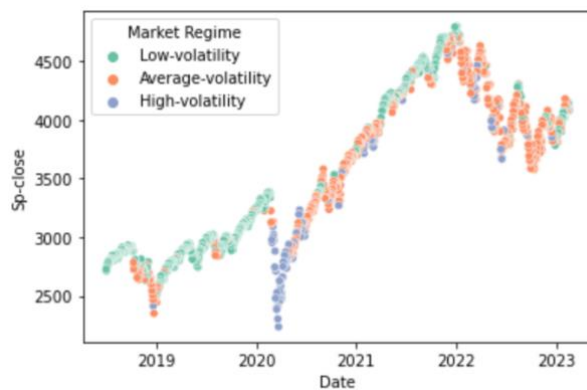
- Տվյալների բարդություն, որքան շատ են տվյալները, այնքան ավելի բարդ է գտնել օրինաչափություններ:
- Շուկայական դինամիկայի փոփոխություն, շուկան փոխվում է ժամանակի ընթացքում, ինչը կարող է ազդել կլաստերավորման ալգորիթմների արդյունավետության վրա: Օրինակ, վերջին 5 տարիներին համաշխարհային տնտեսության մեջ տեղի ունեցած զգալի փոփոխությունները դարձրել են շուկան ավելի անկայուն և դժվար է ճշգրիտ բաժանում կատարել:
- Կլաստերների օպտիմալ քանակ, տվյալների ավելի շատ կետերի դեպքում ավելի դժվար կլինի որոշել կլաստերների օպտիմալ թիվը: Որոշ դեպքերում, չափազանց շատ կամ շատ քիչ կլաստերներ ունենալը կարող է հանգեցնել ավելի քիչ ճշգրիտ կլաստերների:

## Աղյուսակ 6: Տվյալների միջինները

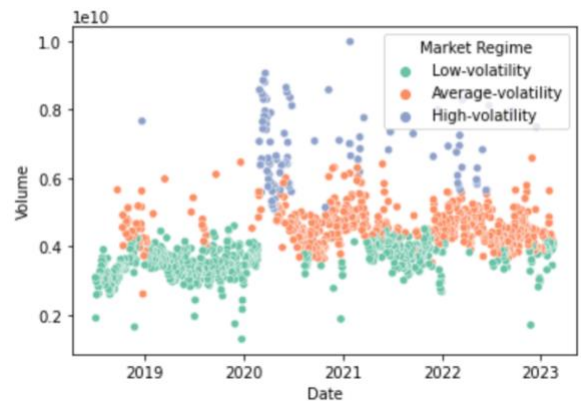
	Sp-close	Volume	Vix-close
kmeans_3			
0	3415.263526	3.550356e+09	16.284491
1	3273.627921	6.825679e+09	38.223168
2	3747.903790	4.619172e+09	24.827984

Ելնելով տվյալների միջիններից կարող ենք առանձնացնել ռեժիմները ըստ VIX-ի:

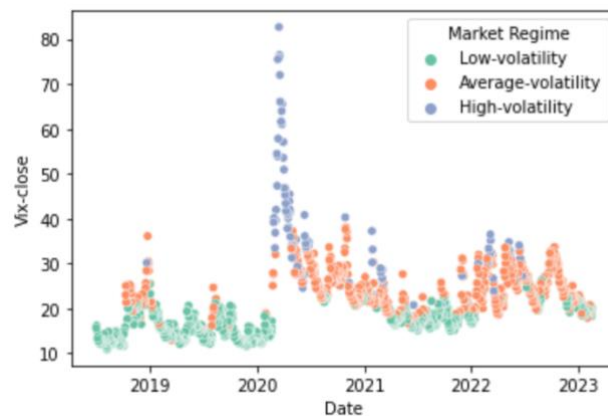
Գծապատկեր 28: S&P close price



Գծապատկեր 29 : S&P Volume



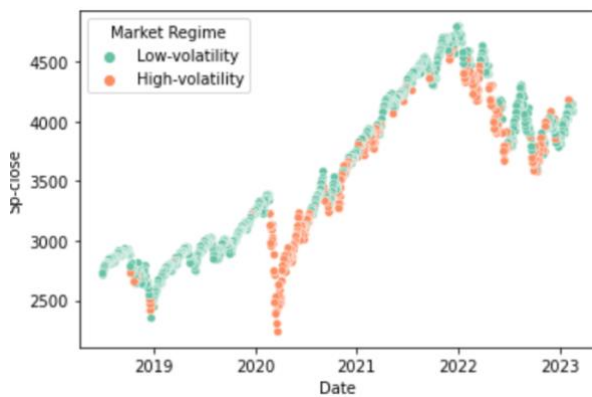
Գծապատկեր 30: VIX index



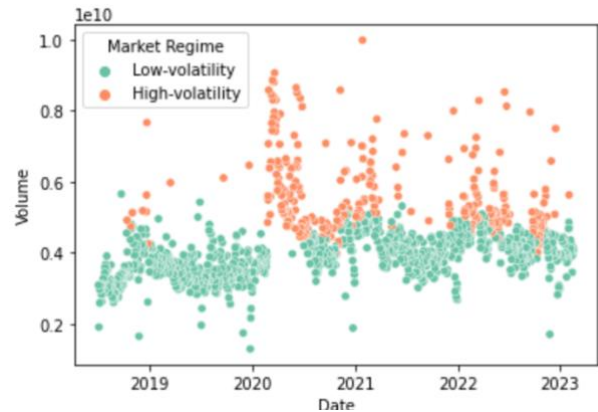
Լոգարիթմական եկամտաբերությունը լավ ազդեցություն է ունեցել կլաստերավորման վրա, այն ավելի լավ է ճշգրտում ռեժիմները:

Այն նաև լավ է աշխատում 2 կլաստերների բաժանելով, երբ օգտագործում ենք log return-ը:

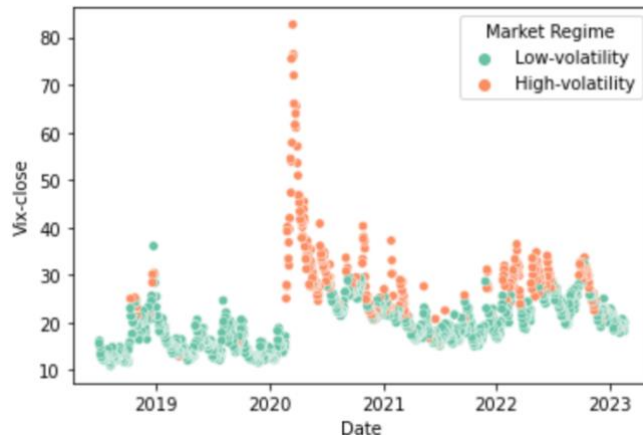
Գծապատկեր 31 : S&P close price



Գծապատկեր 32 : S&P Volume



Գծապատկեր 33: VIX index



Վերը նշված գծապատկերներից երևում է, որ այն կարողացել է առանձնացնել S&P 500-ի փակման գնի տատանումները(Գծապատկեր 31):

### ***GMM(Gaussian Mixture Model)***

Այս մոդելի<sup>2</sup> և Թաքնված Մարկովյան Մոդելների մեջ օգտագործել ենք՝ S&P 500-ի լոգարիթմական եկամտաբերությունը(log return), քանի որ ալգորիթմը դրա միջոցով լավ է գտնում օրինաչափություններ տվյալների միջև, ինչպես նաև առևտրի ծավալը և VIX-ը, ապա համեմատել ստացված արդյունքները, հասկանալու թե որ մուտքային տվյալների դեպքում է ավելի լավ բացահայտում թաքնված վիճակները:



### Գծապատկեր 34: GMM with S&P log return

Number of observations for State 0 : 1123  
Number of observations for State 1 : 129



### Գծապատկեր 35: GMM with S&P log return, Volume)

Number of observations for State 0 : 989  
Number of observations for State 1 : 173



### Գծապատկեր 36: GMM with S&P log return, Volume and VIX

Number of observations for State 0 : 956  
Number of observations for State 1 : 200



### Գծապատկեր 37: GMM with S&P log return, VIX

Number of observations for State 0 : 959  
Number of observations for State 1 : 203



Վերը նշված 4 գծապատկերներն էլ բացահայտում են 2018-ի անկումը, 2020-ի COVID 19-ը և 2022-ից հետո եկող անկայունության շրջանը: Գծապատկեր 36-ը և 37-ը այդքան էլ չեն տարբերվում իրարից, այնուամենայնիվ դժվար է տարանջատել թե որ feature-ների դեպքում է մոդելը ավելի լավ նկարագրում շուկայական ռեժիմները, բայց ի տարբերություն K-means-ի, GMM-ն ավելի լավ է բացահայտում թաքնված վիճակները:

### Շուկայի ռեժիմների հայտնաբերում Թաքնված Մարկովյան Մոդելներ միջոցով

Տեսնենք թե ինչպես է Թաքնված Մարկովյան Մոդելը կարողանում բացահայտել թաքնված վիճակները: Մենք կրկին կդիտարկենք մի քանի մուտքային տվյալ, ապա կհամեմատենք դրանք:

Գծապատկեր 38: HMM with S&P log return

Number of observations for State 0 : 104  
Number of observations for State 1 : 1148



Գծապատկեր 39: HMM with S&P log return, VIX

Number of observations for State 0 : 494  
Number of observations for State 1 : 668



Գծապատկեր 40: HMM with S&P log return, Volume

Number of observations for State 0 : 406  
Number of observations for State 1 : 756



#### Գծապատկեր 41: HMM with S&P log return, Volume, VIX

Number of observations for State 0 : 642  
Number of observations for State 1 : 520



Վերը նշված 4 գծապատկերներից վիզուալ տեսանկյունից Գծապատկեր 38-ը բացահայտում է 2018-ի անկումը, 2020-ի COVID 19-ը և 2022-ից հետո եկող անկայունության շրջանը: Գծապատկեր 39-ը 40-ը և 41-ի դեպքում, դժվար է տարանջատել թե որ մոդելն է ավելի լավ նկարագրում շուկայական ռեժիմները, այնուամենայնիվ հարկավոր է կառուցել առևտրային ստրատեգիաներ հիմնված այս նմուշների վրա, որպեսզի կարողանանք նկարագրել յուրաքանչյուր մոդելի արդյունավետությունը:

Հետազոտության արդյուքները կարող եք տեսնել գրականության ցանկին կից Github<sup>9</sup>-ի կայքում:

## ԵԶՐԱԿԱՑՈՒԹՅՈՒՆ

Այս աշխատության մեջ կիրառեցինք մի քանի վիճակագրական և մեքենայական ուսուցման մոդելներ շուկայի ռեժիմների հայտնաբերման համար: Պարզեցինք թե դրանցից որոնք են ավելի արդյունավետ կերպով բացահայտում շուկայում թաքնված վիճակները:

Սկզբում ներկայացրեցինք կլաստերավորման երկու ալգորիթմ՝ K-means-ը և Գաուսյան Խառնուրդի Մոդելը: Տեսանք, որ K-means-ը այդքան էլ արդյունավետ չէ շուկայի ռեժիմների հայտնաբերման տեսանկյունից, մինչդեռ Գաուսյան Խառնուրդի Մոդելը և Թաքնված Մարկովյան Մոդելները կարող են արժեքավոր պատկերացումներ տալ ֆոնդային շուկայի վարքագծի վերաբերյալ և օգնել ներդրողներին ավելի տեղեկացված որոշումներ կայացնել:

Նշենք, որ վերլուծությունները կատարվել են նմուշի տեսքով՝ ազդելով արդյունքների կայունության վրա: Մոդելը կիրառական դարձնելու համար, անհրաժեշտ է իրականացնել նմուշից դուրս փորձարկում և կառուցել ներդրումային ռազմավարություններ այդ արդյունքների հիման վրա:

Հետազոտության հիմնական նպատակը ոչ թե իրական աշխարհի առևտրի մոդել տրամադրելն էր, այլ ցուցադրել շուկայական ռեժիմների հայտնաբերման տարբեր մեթոդներ, որոնք կարող են օգտագործվել առաջադեմ ներդրումային ռազմավարություններ կառուցելիս:

---

<sup>1</sup> <https://www.quantstart.com/articles/hidden-markov-models-for-regime-detection-using-r/>

<sup>2</sup> <https://developers.refinitiv.com/en/article-catalog/article/market-regime-detection>

<sup>3</sup> [https://hy.wikipedia.org/wiki/Մեթինայական\\_ուսուցում](https://hy.wikipedia.org/wiki/Մեթինայական_ուսուցում)

<sup>4</sup> [https://en.wikipedia.org/wiki/Cluster\\_analysis](https://en.wikipedia.org/wiki/Cluster_analysis)

<sup>5</sup> Kevin P. Murphy. Machine learning: a probabilistic perspective, 2012.

<sup>6</sup> <https://neptune.ai/blog/clustering-algorithms-~-:text=Clustering algorithms are used to,specific user and the scenario.>

<sup>7</sup> Daniel Jurafsky & James H. Martin, Speech and Language Processing. Copyright © 2023. All rights reserved. Draft of January 7, 2023.

<sup>8</sup> <https://vitalflux.com/hidden-markov-models-concepts-explained-with-examples/>

GitHub URL`

<sup>9</sup> <https://github.com/mhovsepy/Identification-of-Market-Regimes.git>