



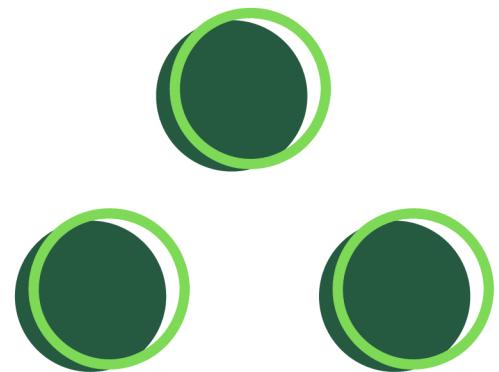
Principal Component Analysis (PCA)

Գլխավոր Բաղադրիչների Վերլուծություն



ԳԼԽԱՎՈՐ ԲԱՂԱԴՐԻՉՆԵՐԻ ՎԵՐԼՈՒԾՈՒԹՅՈՒՆ (PCA)

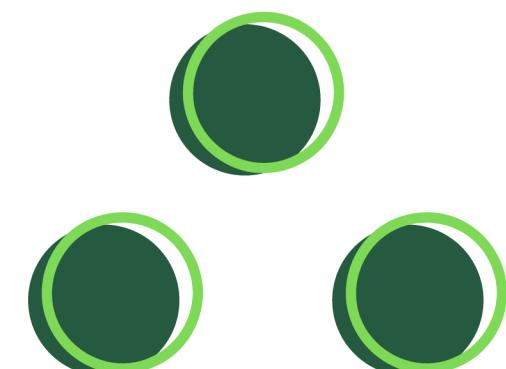
Մեքենայական ուսուցման մեթոդ, որն օգտագործվում է բազմաթիվ հատկանիշներով տվյալների չափողականության նվազեցման համար, միաժամանակ պահպանելով դրանց մեջ առկա կարևոր օրինաչափությունները և ինֆորմացիան:



PCA-ն գծային մոդել է, որը կառուցում է նոր կոորդինատային համակարգ՝ օգտագործելով սկզբնական տվյալների գծային կոմբինացիաները: Նպատակը այն է, որ նոր համակարգում տվյալների պրոյեկցիաները առանցքների վրա ունենան առավելագույն դիսպերսիա (variance):

Կիրառություններ

- Տվյալների վիզուալիզացիա
- Տվյալների թեթևացում (գերուսուցման հավանականության նվազեցում)
- Տվյալների ֆիլտրում (Անօգտակար հայտանիշների հեռացում)
- Ուսուցման ալգորիթմի արագացում



Դիսպերսիա (Variance)

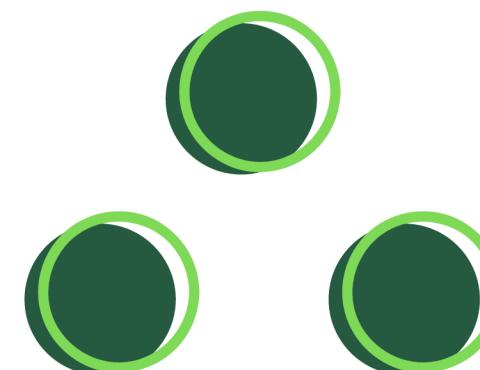
Դիսպերսիան ցույց է տալիս, թե որքանով են տվյալները ցոված իրենց միջինից:

$$\text{Var}(X_i) = \frac{1}{n-1} \sum_{k=1}^n (x_{ki} - \bar{X}_i)^2$$

Կովարիացիոն մատրից (Covariance Matrix)

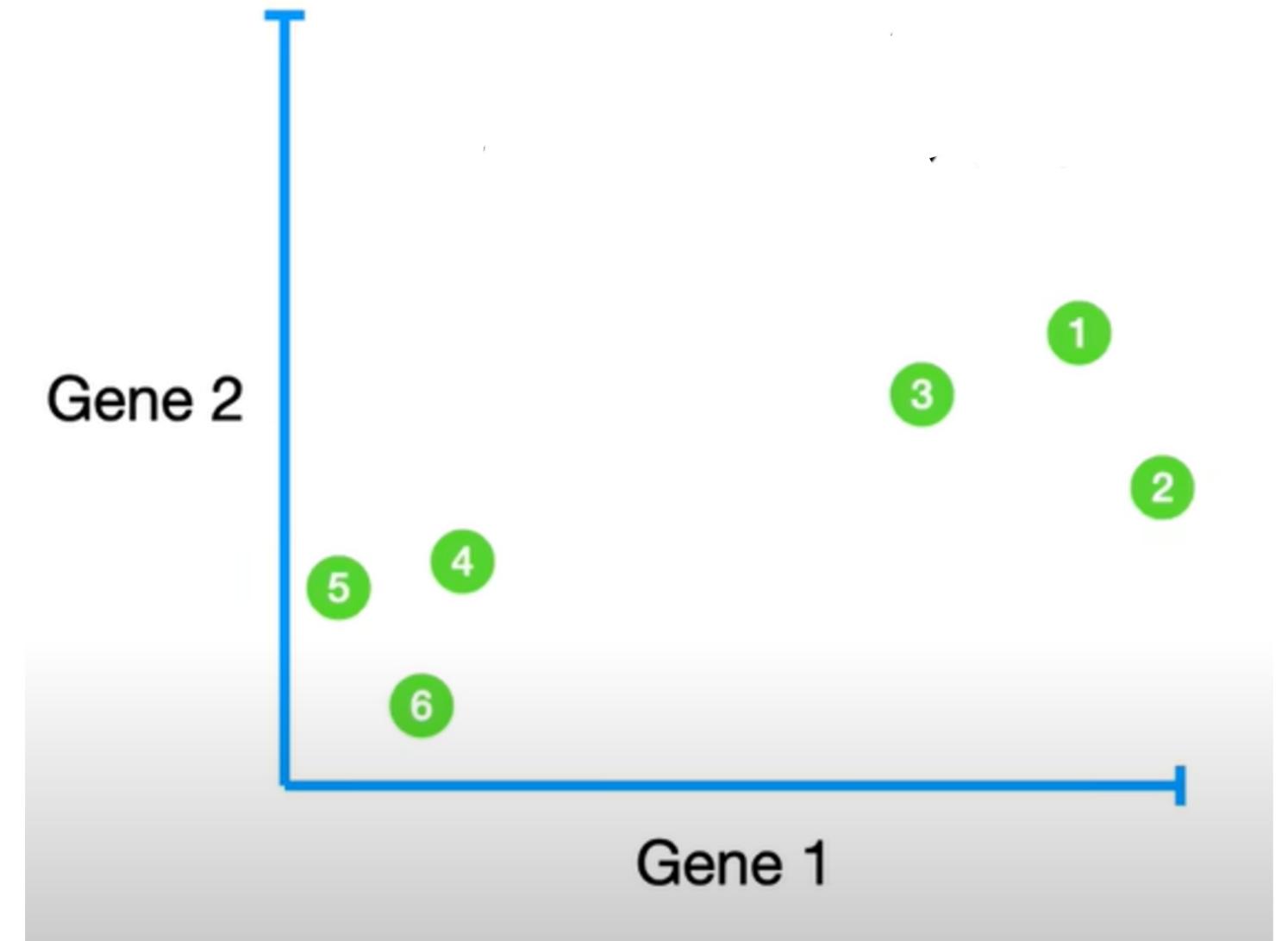
Իրենից ներկայացնում է քառակուսի մատրից որի (i,j) վանդակում գրվում է (X_i, X_j) փոփոխականների կովարիացիան(թե ինչքանով են նրանք նման անում):

$$\text{Cov}(X_i, X_j) = \frac{1}{n-1} \sum_{k=1}^n (x_{ki} - \bar{X}_i)(x_{kj} - \bar{X}_j)$$

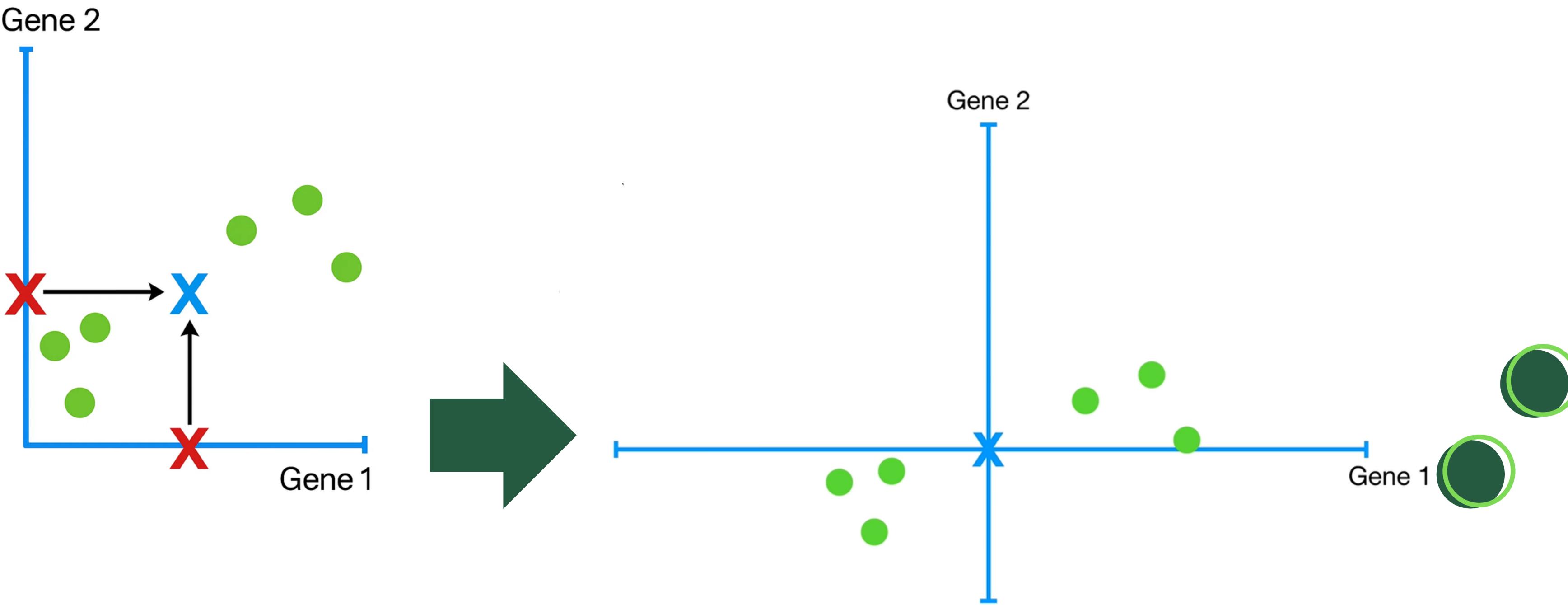


Պրոցեսի պարզ նկարագրություն օրինակով

	Mouse 1	Mouse 2	Mouse 3	Mouse 4	Mouse 5	Mouse 6
Gene 1	10	11	8	3	1	2
Gene 2	6	4	5	3	2.8	1

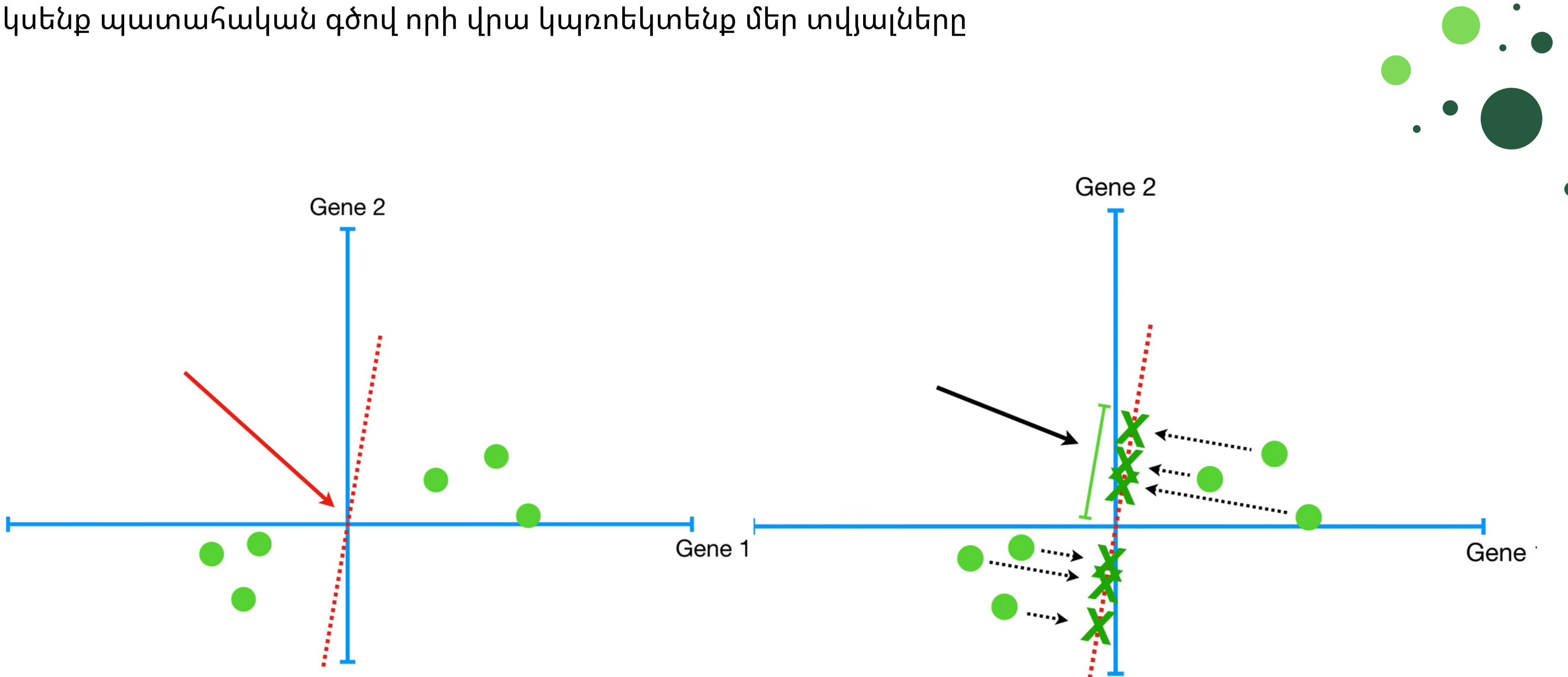


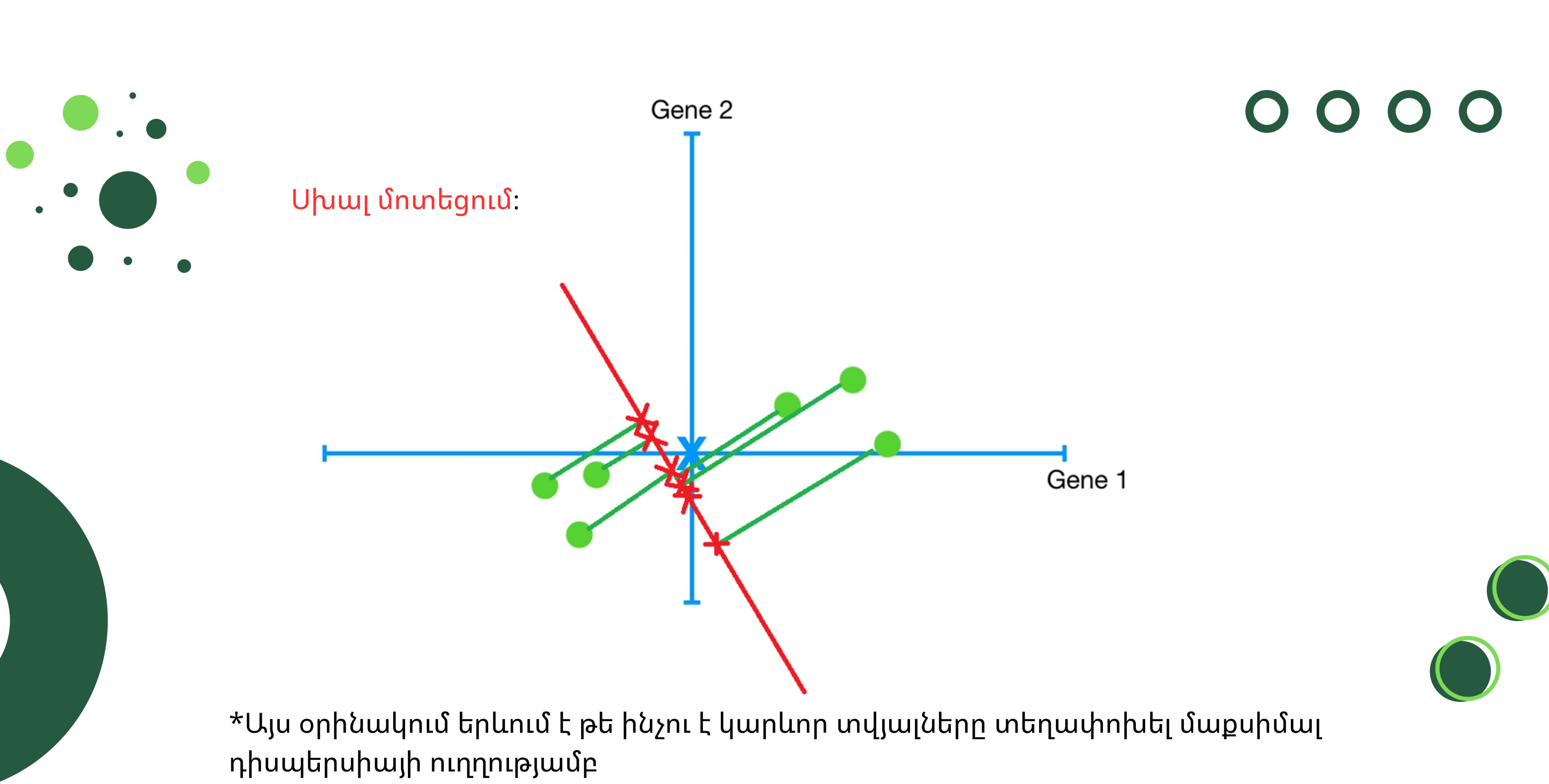
Տվյալները տեղաշարժենք՝ կենտրոնը բերելով (0,0) (centering)



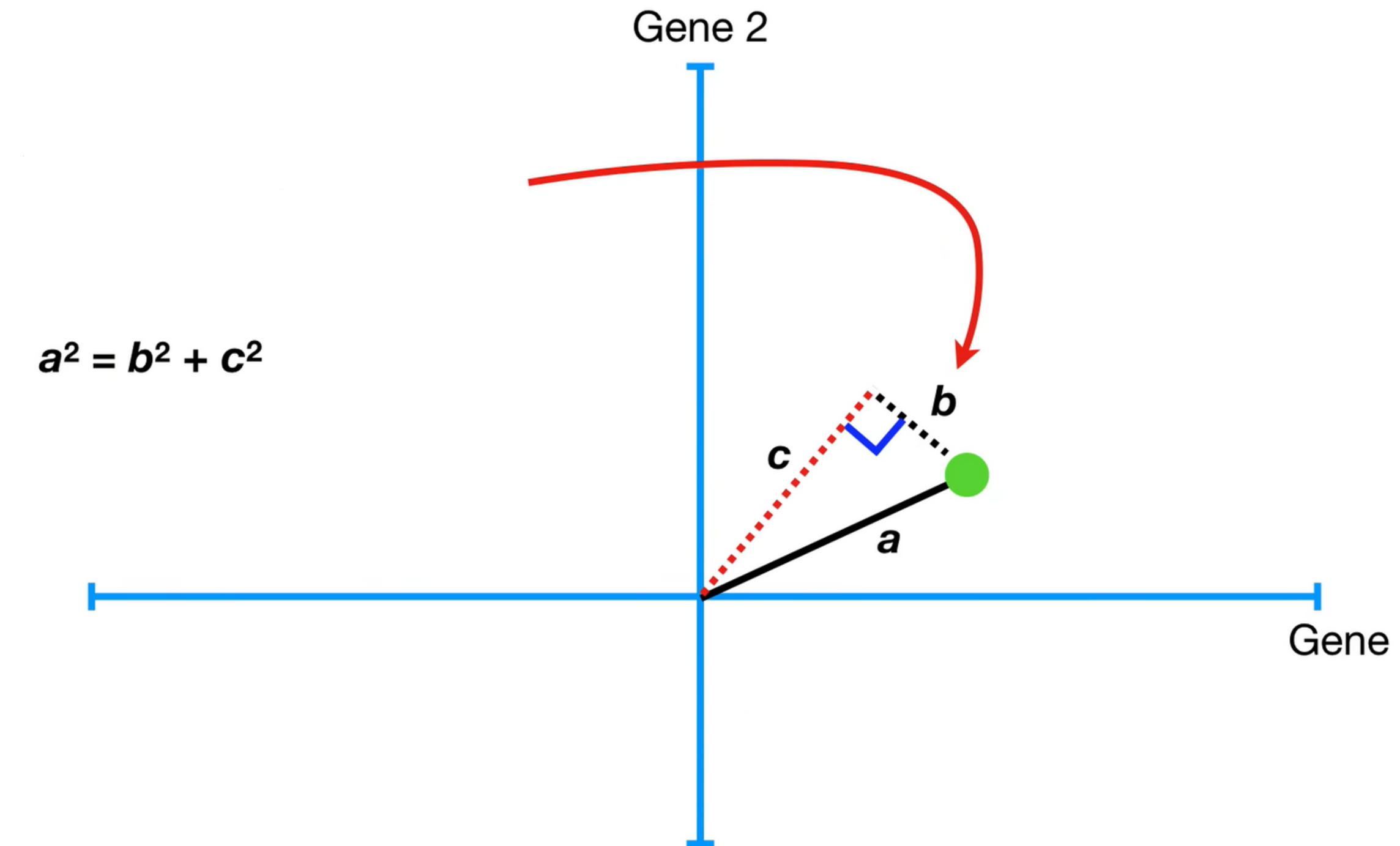
ՄԵՐ ՆԱՎԱՏԱԿՆ Է ՄԵՐ ՏՎՅԱԼՆԵՐԸ ԲԵՐԵԼ 1 ՀԱՎԻԱՆԻ ՏԱՐԱԾՈՒԹՅՈՒՆ , ՀԿՈՐՉՈՆԵԼՈՎ ՉԱՄ ԻՆՖՈՐՄԱՑԻԱ

*Ակսենք պատահական գծով որի վրա կառուեկտենք մեր տվյալները



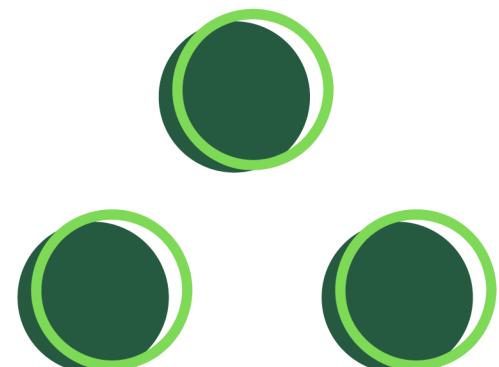
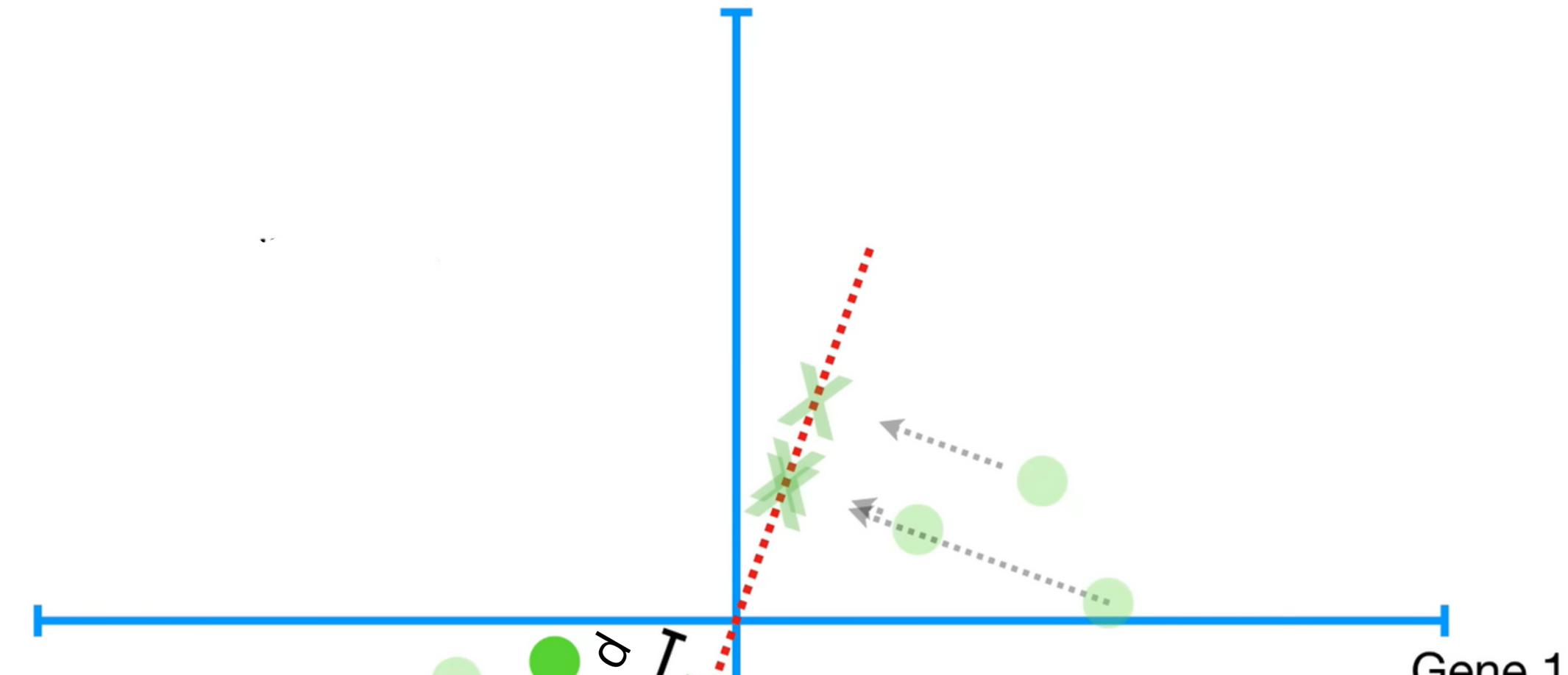


Մաքսիմիզացնել դիսպերսին = մինիմիզացնել կետի և ուղղի հեռավորությունը

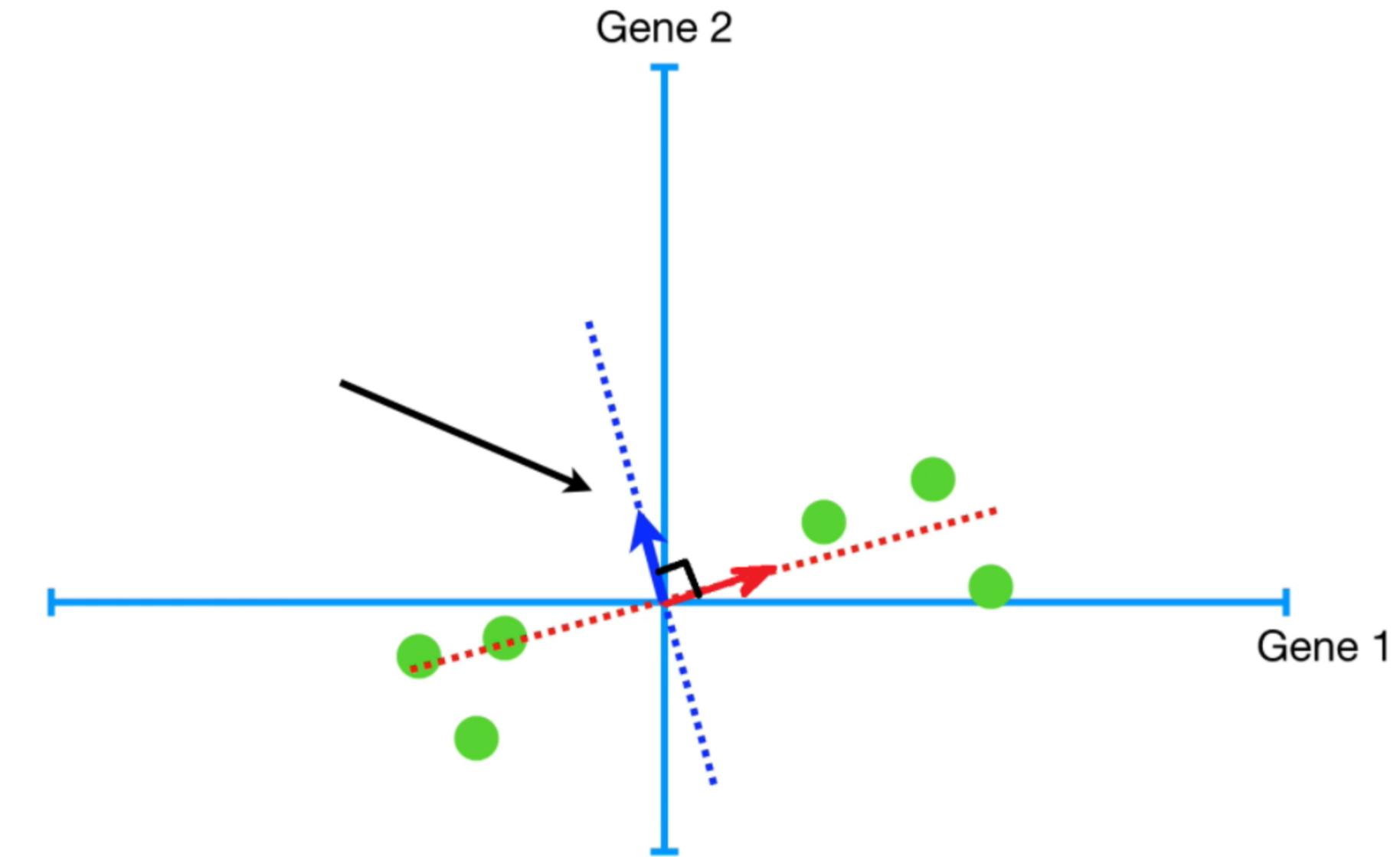
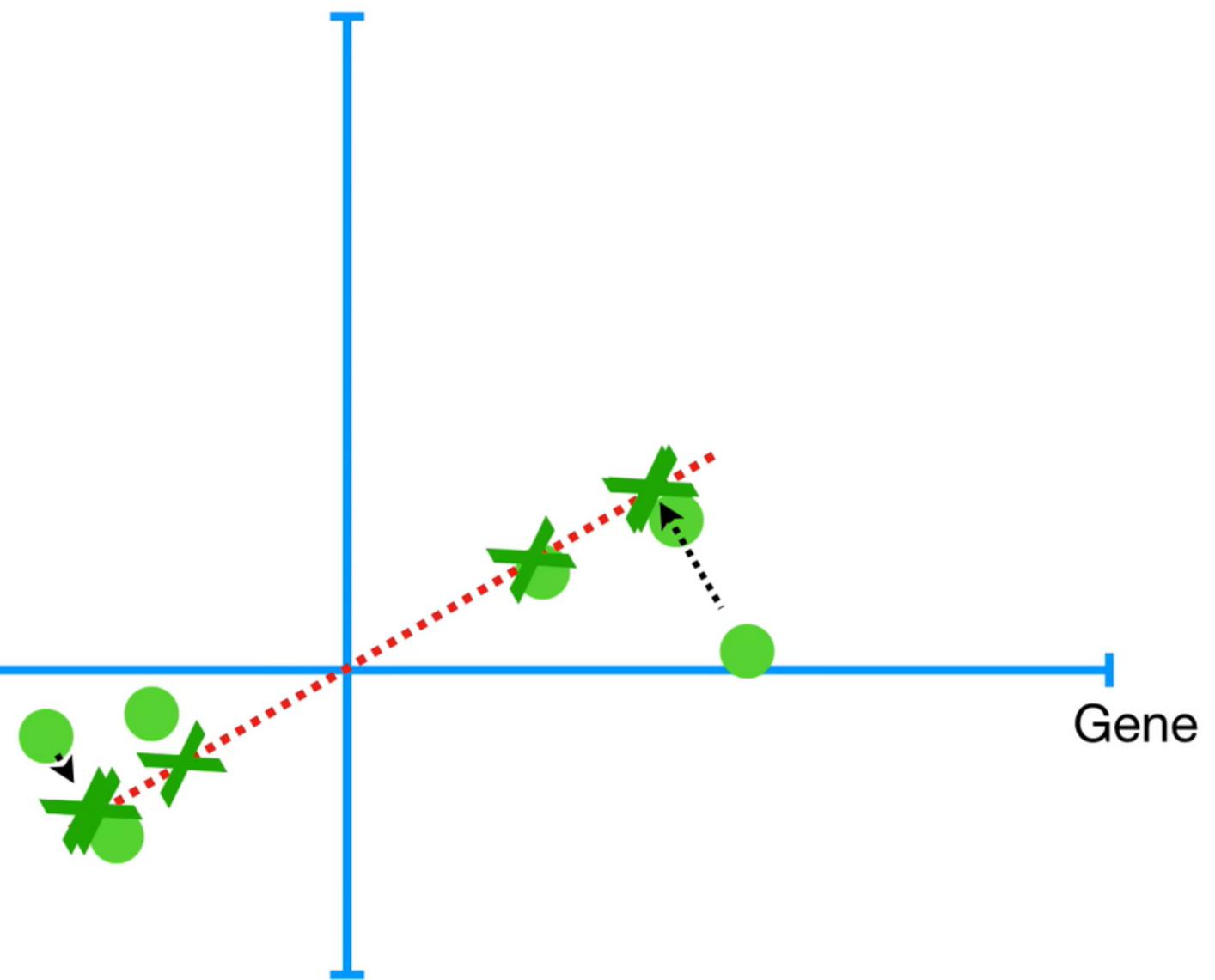


Տեղափոխենք ուղիղը այնպէս որ ամեն քայլում մեծացնենք դիսպերսիան

$$d_1^2 + d_2^2 + d_3^2 + d_4^2 + d_5^2 + d_6^2 = \text{variance}$$

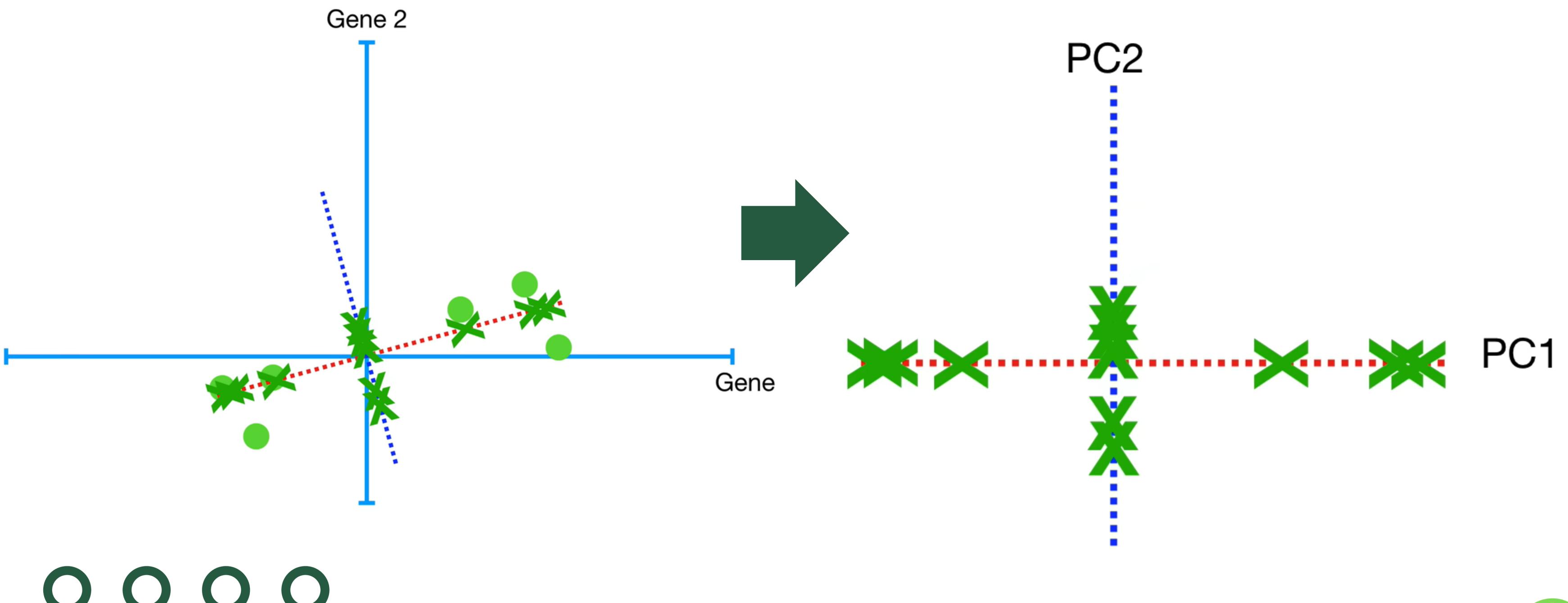


- Արդյունքում կստանանք ձախ նկարի կարմիր ուղիղը որի անունն է **PC1** (**Principal component 1**) այն կհանդիսանա նոր կոորդինատային համակարգի առաջին պարամետրը
- Եթե ցանկանում ենք 2րդ պարամետր էլ ստանալ, աջ նկարում միանգամից գծված է **PC2** ը որը պետք է լինի ուղղահայաց PC1 ին(սա հետևում է մաթեմատիկական դուրս բերումից որին կանդրադառնանք)

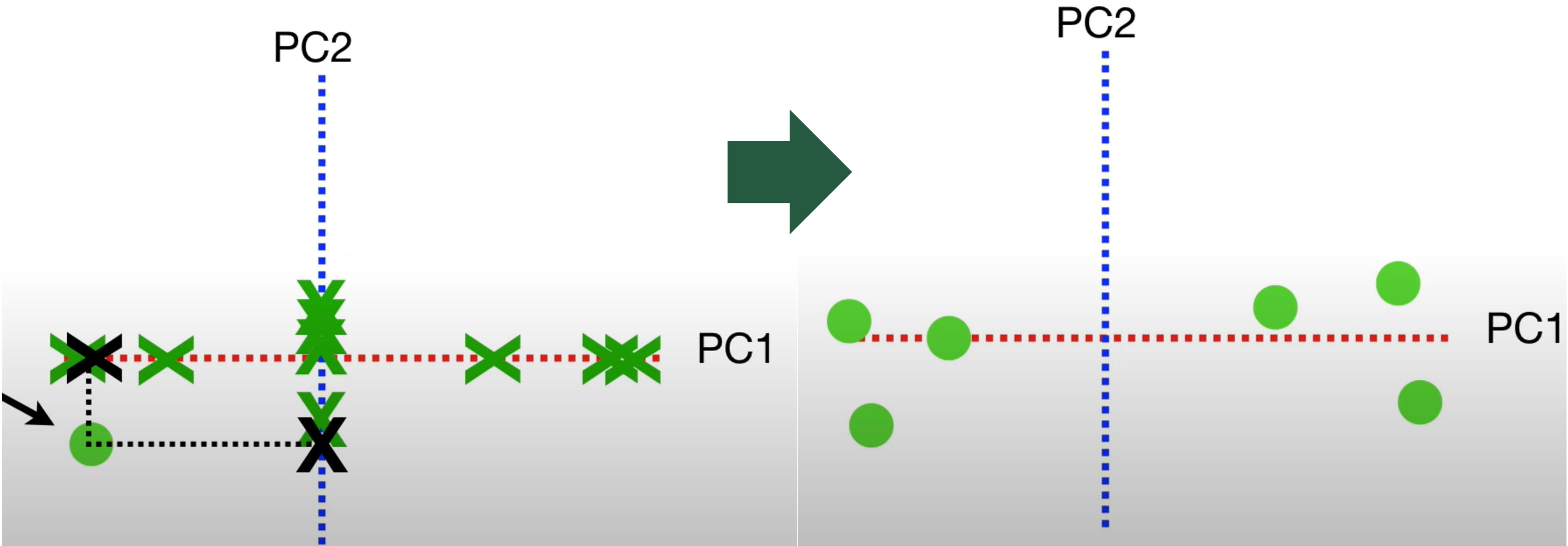


Պռուեկտենք կետերը նոր ուղիղների վրա

Պտտենք՝ ստանալով նոր կոորդինատային համակարգը

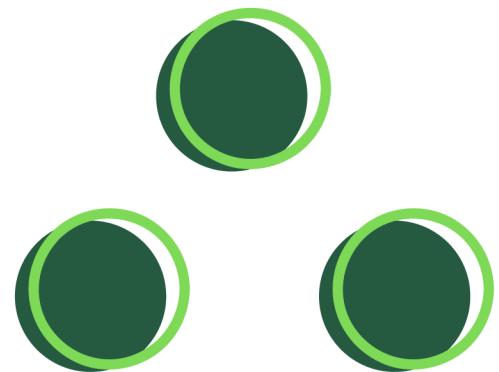


ԱՐԴՅՈՒՆՔ



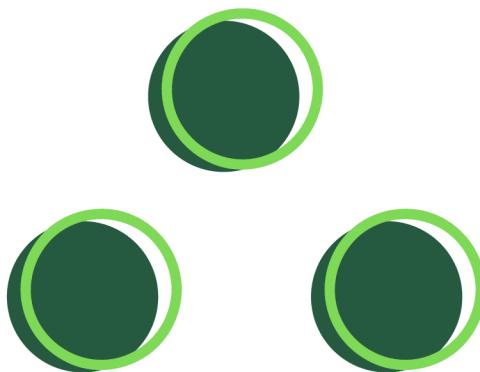
ՄՆՈՒՄ Է ՀԱՍԿԱՆԱԼ ԹԵ ՌԵՆԳ ԸՆՏՐԵԼ ՄԱՔՍԻՄԱԼ ԴԻՍԱԲԵՐՍԻԱՅԻ ՈՒՂՂՈՒԹՅՈՒՆը

- Գծային հանրհաշվից հայնտի արդյունքներից օգտվելով ապացուցվում է որ Դիսաբերսիայի մաքսիմալ աճի ուղղությունը(PCA1) համընկնում է մեր տվյալներին համապատասխանող կովարիացիոն մատրիցի այն սեփական վեկտորին որի սեփական արժեքը ամենամեծն է:
- Մնացած Principal Component - ները համընկնում են համապատասխանաբար այն սեփական վեկտորի հետ որի սեփական արժեքը մեծությամբ հաջորդն է:



Քայլերի հաջորդականություն (և գլխավոր կոմպոնենտ)

- Տվյալների կենտրոնացում $X_{\text{centered}} = X - \mu$
- Կովարիացիոն մատրիցի հաշվարկ $\text{Cov}(X) = \frac{1}{n-1} X_{\text{centered}}^T X_{\text{centered}}$
- Մատրիցի սեփական ֆունկցիաների և արժեքների հաշվարկ $\text{Cov}(X) \cdot v = \lambda \cdot v$
- Սորտավորում ենք սեփական վեկտորները ըստ սեփական արժեքների և ընտրում առաջին և սեփական վեկտորները որպես նոր պարամետրերի ուղղություններ
- Տվյալները պրոյեկտում ենք նոր տարածության վրա





ԻԿՄ Յ-րդ կուրս



ԱՆՆԱ ԿԱՐԱՊԵՏՅԱՆ, ԱՐՄԱՆ ՍԱՐԳՍՅԱՆ

