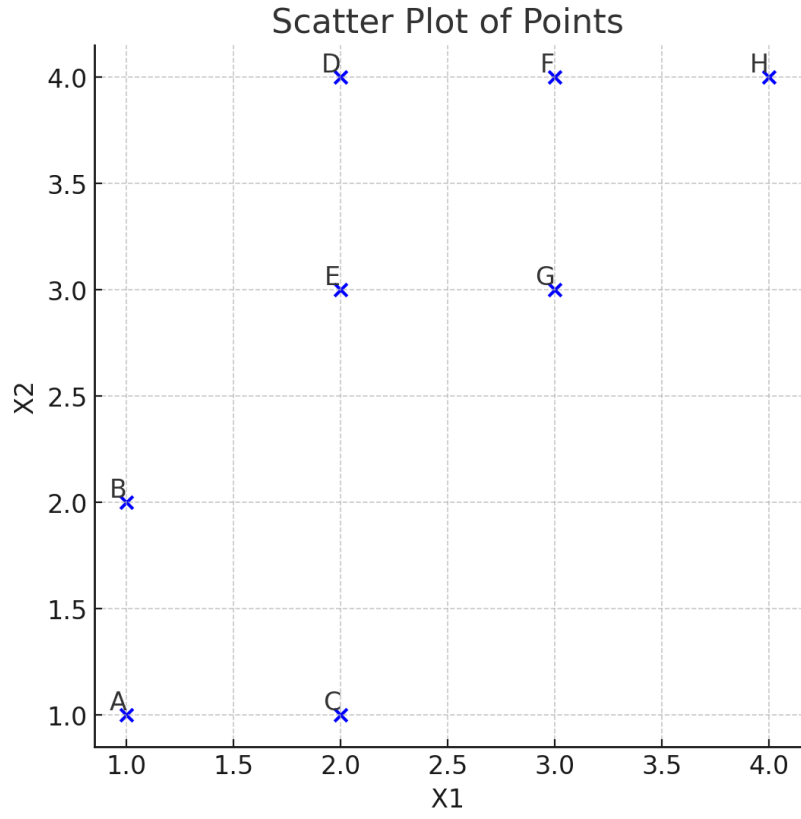


## K-Means Alıştırması

Veriler			
Nokta	X1	X2	Renk
A	1	1	
B	1	2	
C	2	1	
D	2	4	
E	2	3	
F	3	4	
G	3	3	
H	4	4	

Kırmızı M.		
Mavi M.		

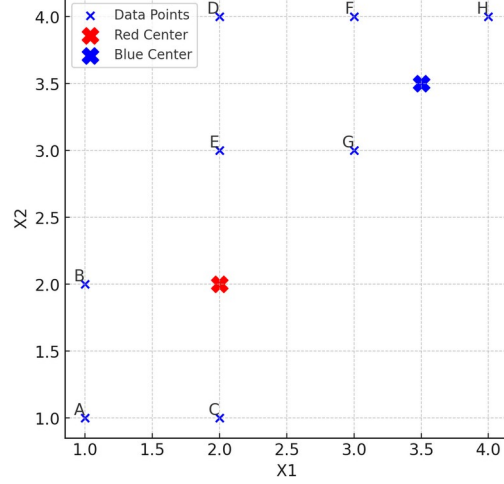


K-Means algoritması, başlangıçta rastgele seçilmiş merkezler kümesi ile çalışmaya başlar.

Adım	1
Kırmızı M.	2 2
Mavi M.	3,5 3,5

Veriler					
Nokta	X1	X2	$D_K$	$D_M$	Renk
A	1	1	2	5	Red
B	1	2	1	4	Red
C	2	1	1	4	Red
D	2	4	2	2	Red
E	2	3	1	2	Red
F	3	4	3	1	Blue
G	3	3	2	1	Blue
H	4	4	4	1	Blue

Initial Cluster Centers for K-Means (Manhattan Distance)



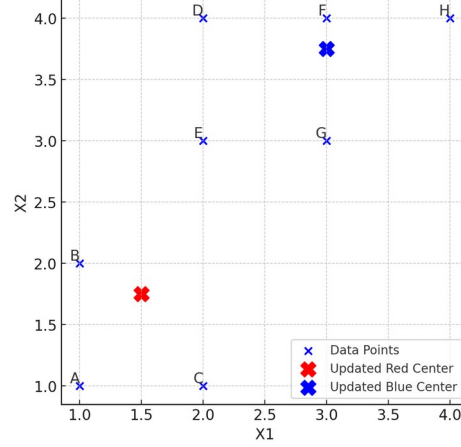
## K-Means Alıştırması

Sonraki yinelemelerde, merkezler, bir önceki yinelemede yapılan sınıflandırmaya (kümelemeye) göre yeniden hesaplanır.

Adım	2
Kırmızı M.	1,5 1,8
Mavi M.	3 3,8

Veriler					
Nokta	X1	X2	D <sub>K</sub>	D <sub>M</sub>	Renk
A	1	1	1,25	4,75	Red
B	1	2	2,75	3,75	Red
C	2	1	1,25	3,75	Red
D	2	4	2,75	1,25	Blue
E	2	3	1,75	1,75	Red
F	3	4	3,75	0,25	Blue
G	3	3	2,75	0,75	Blue
H	4	4	4,75	1,25	Blue

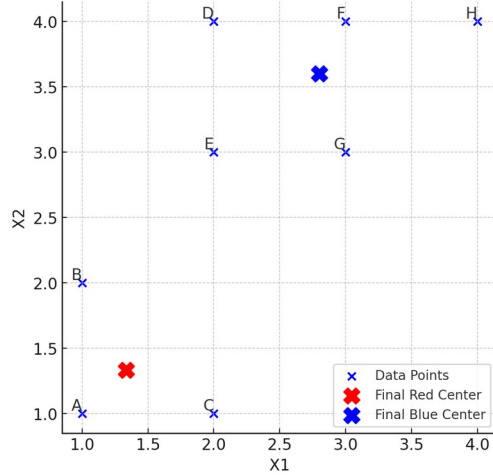
Updated Cluster Centers for K-Means (Manhattan Distance)



Adım	3
Kırmızı M.	1,3 1,3
Mavi M.	2,8 3,6

Veriler					
Nokta	X1	X2	D <sub>K</sub>	D <sub>M</sub>	Renk
A	1	1	0,66	4,4	Red
B	1	2	1	3,4	Red
C	2	1	1	3,4	Red
D	2	4	3,33	1,2	Blue
E	2	3	2,33	1,4	Blue
F	3	4	4,33	0,6	Blue
G	3	3	3,33	0,8	Blue
H	4	4	5,33	1,6	Blue

Final Cluster Centers for K-Means (Manhattan Distance)



Adım	4
Kırmızı M.	1,3 1,3
Mavi M.	2,8 3,6

Merkezler kararlı hale geldiğinde, yani artık önceki yinelemeye göre değişmiyorsa, mevcut sınıflandırmamız da kararlı (stabil) hale gelmiş demektir.

### Notlar:

- 1- Ondalık sayılar bilgisayarlarda tam olarak saklanmadığı için, belirli bir kayan nokta formatında depolanırlar. Bu yüzden bazen yuvarlama hataları, merkez hesaplamalarının tam olarak sabitlenmeden etrafında hafifçe hareket etmesine neden olabilir. Bu durumu kontrol altına almak için, küçük bir tolerans değeri tanımlayarak merkez hesaplamalarındaki değişimin bu tolerans sınırları içerisinde olup
- 2- Alternatif bir yöntem ise, \*\*sınıflandırmalarda (küme atamalarında)\*\* herhangi bir gerçek değişiklik olup olmadığını kontrol etmektir.

## K-Means Alıştırması

### Soru:

Eğer merkezleri sabit tutup, ardından yeni noktalar eklemeye başlarsak (merkezleri güncellemeden K-Means tahmini), “concept drift”i yönetebilir miyiz?

**Ani Kayma (Sudden Drift)** Veri dağılımında ani ve hızlı bir değişim meydana gelir, önceki desenler anında geçersiz hale gelir. Örnek: Büyük bir olay sonrası tüketici davranışlarının aniden değişmesi.

**Kademeli Kayma (Gradual Drift)** Veri dağılımı zamanla yavaşça kayar ve eski ile yeni kavramlar bir süre boyunca bir arada var olabilir. Örnek: Yeni bir ürün özelliğinin yavaş yavaş popülerlik kazanması, eskisinin ise hâlâ bir miktar kullanılmaya devam etmesi.

**Artımlı Kayma (Incremental Drift)** Zaman içinde biriken küçük ve sürekli değişimler, veri dağılımında önemli bir kaymaya yol açar. Örnek: Bir kullanıcının tercihlerinin uzun vadeli davranış eğilimlerine bağlı olarak yavaşça evrilmesi.

**Tekrarlayan Kayma (Recurring Drift)** Daha önce görülen veri dağılımları periyodik olarak geri döner. Örnek: Kara Cuma veya Yeni Yıl gibi dönemlerde periyodik olarak tekrar eden perakende alışveriş alışkanlıkları.

“Kayma (drift)” olup olmadığını kontrol etmek için kullanılan yöntemler:

Kolmogorov–Smirnov (K–S) Test

Jensen–Shannon Divergence (JSD)

Population Stability Index (PSI)