

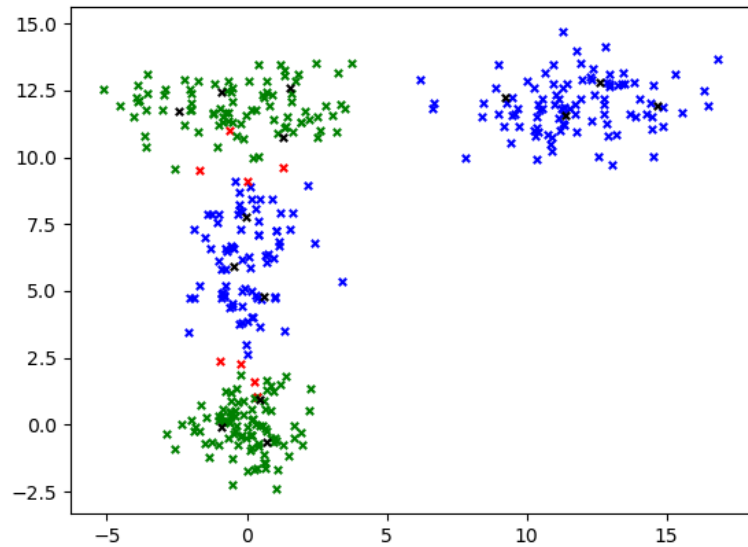
# یافتن مقدار بهینه ی $m$ :

بدین منظور مجموع دقت داده های آموزشی و تست در بازه ی  $m:[start=2, end=20, step=2]$  مورد بررسی قرار گرفت و بالاترین مجموع انتخاب شد:

بهترین مقدار  $m$  برای تست کیس 2clstrain1200 : 14

دقت داده های آموزشی: 0.9857

دقت داده های تست: 0.977

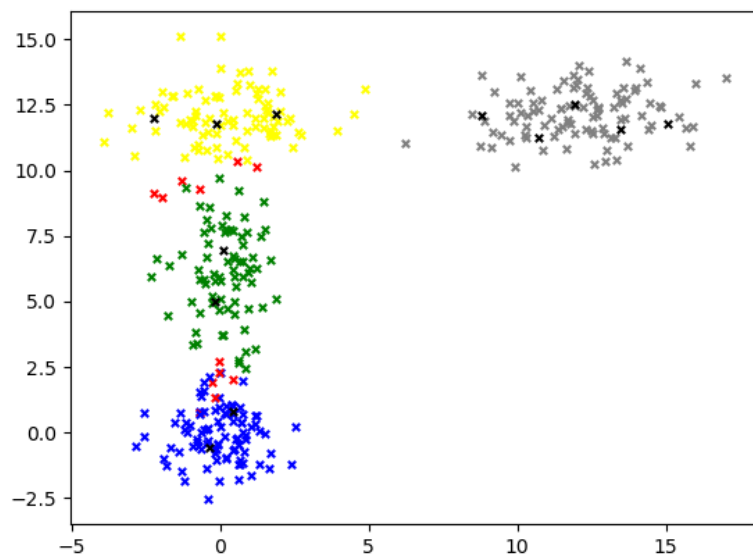


❖ نقاط قرمز داده هایی هستند که برچسب شان نادرست تشخیص داده شد و نقاط قرمز مراکز دسته ها می باشند.

بهترین مقدار  $m$  برای تست کیس 4clstrain1200 : 12

دقت داده های آموزشی: 0.975

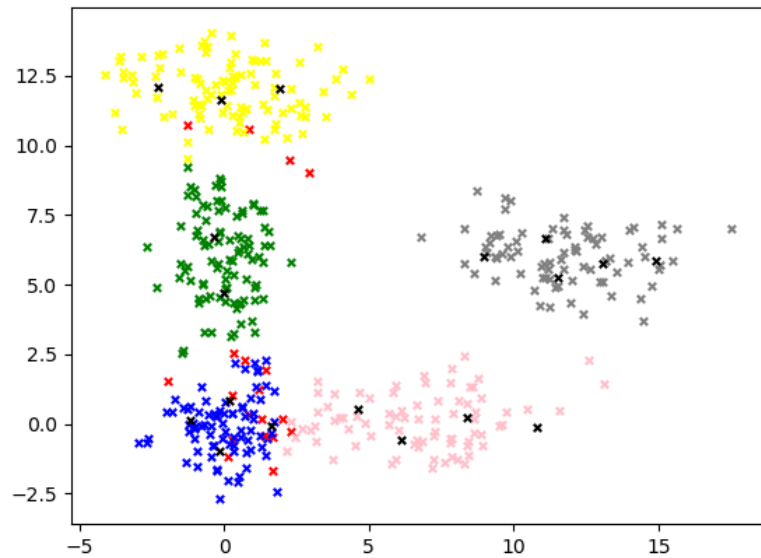
دقت داده های تست: 0.966



بهترین مقدار  $m$  برای تست کیس 5clstrain1500 : 18

دقت داده های آموزشی: 0.9609

دقت داده های تست: 0.9577



در تمام این حالات،  $r = 0.1$  در نظر گرفته شده است.

به طور کلی به نظر می رسد بهترین مقدار برای  $r$  بین  $r = 0.1$  تا  $r = 9$  باشد و در این بازه بالاترین میزان دقت مشاهده شده است. اما در میان مقادیر مختلف داخل همین بازه، تفاوت چندانی در جواب های نهایی مشاهده نشد.

در حالتی که  $m = 40$  و  $r = 0.1$  باشد:

▪ **2clstrain1200 :**

دقت داده های آموزشی: 0.9821

دقت داده های تست: 0.988

▪ **4clstrain1200 :**

دقت داده های آموزشی: 0.9797

دقت داده های تست: 0.9805

▪ **5clstrain1500 :**

دقت داده های آموزشی: 0.9647

دقت داده های تست: 0.9577

در حالتی که  $m = 3$  و  $r = 1$  باشد:

▪ **2clstrain1200 :**

دقت داده های آموزشی: 0.7523

دقت داده های تست: 0.7416

▪ **4clstrain1200 :**

دقت داده های آموزشی: 0.7476

دقت داده های تست: 0.755

▪ **5clstrain1500 :**

دقت داده های آموزشی: 0.6009

دقت داده های تست: 0.5977

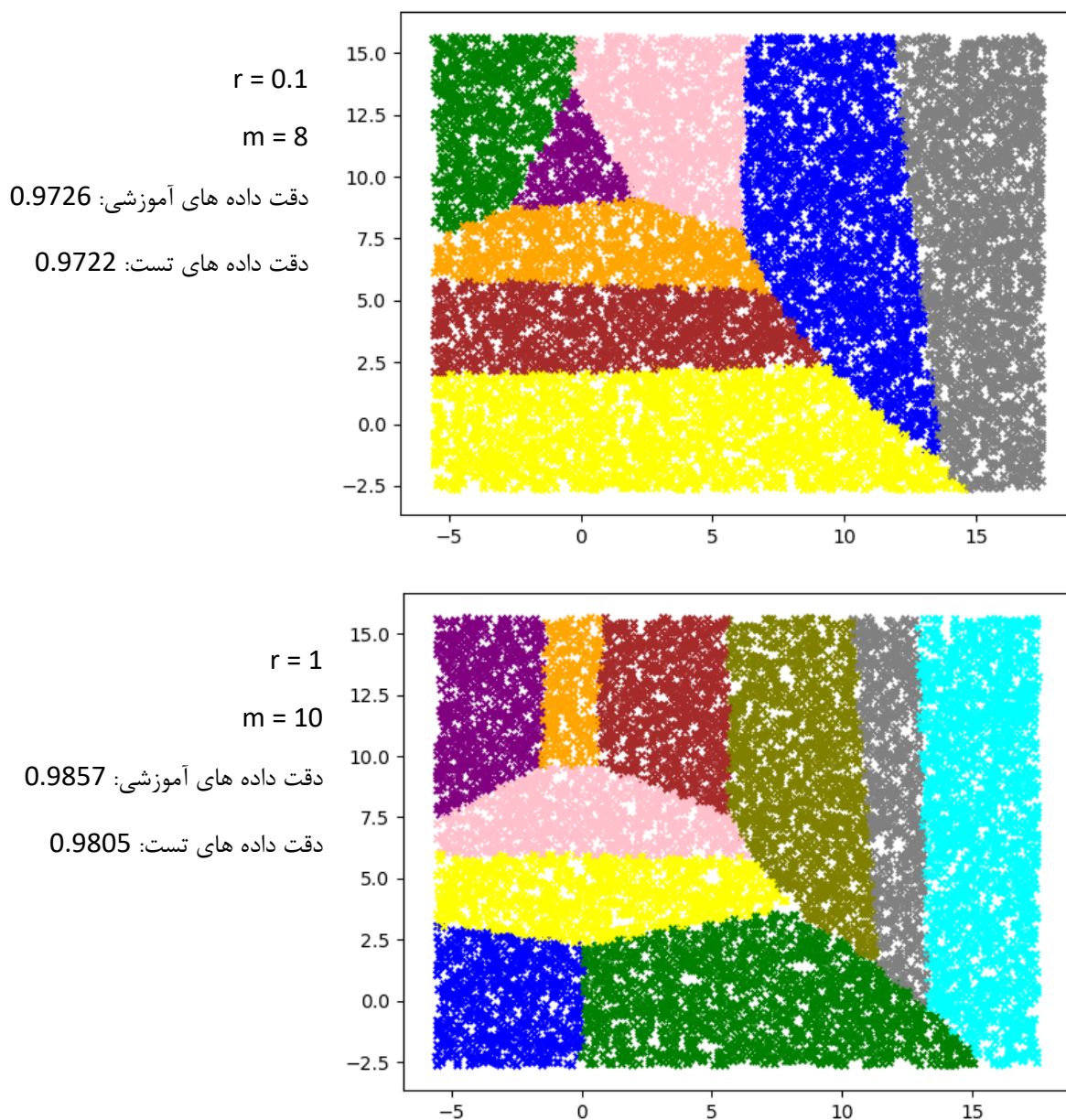
## رسم نتایج مرز بندی خوشه ها:

در این قسمت ، بازه بندی برای یافتن بهترین  $m$  بدین صورت بوده است:  $m:[start=2, end=10, step=2]$

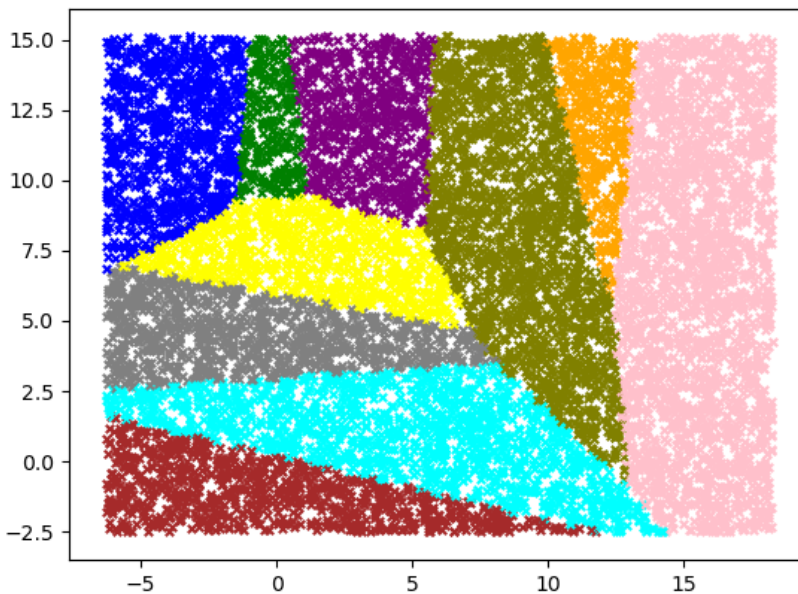
دلیل تغییر بازه ی  $m$  این است که دقت الگوریتم برای مقادیر  $11 < m < 20$  تفاوت چندانی نمی کند و همچنین کم کردن تعداد  $m$  های مورد بررسی برای تعیین  $m$  بهینه، باعث کاهش زمان اجرای برنامه خواهد شد.

هر بار 10,000 داده ی تصادفی تولید شده و با داشتن مرکز خوشه ها از محاسبات قبل، ایندکس بیشترین تعلق به عنوان برچسب هر داده در نظر گرفته شده است.

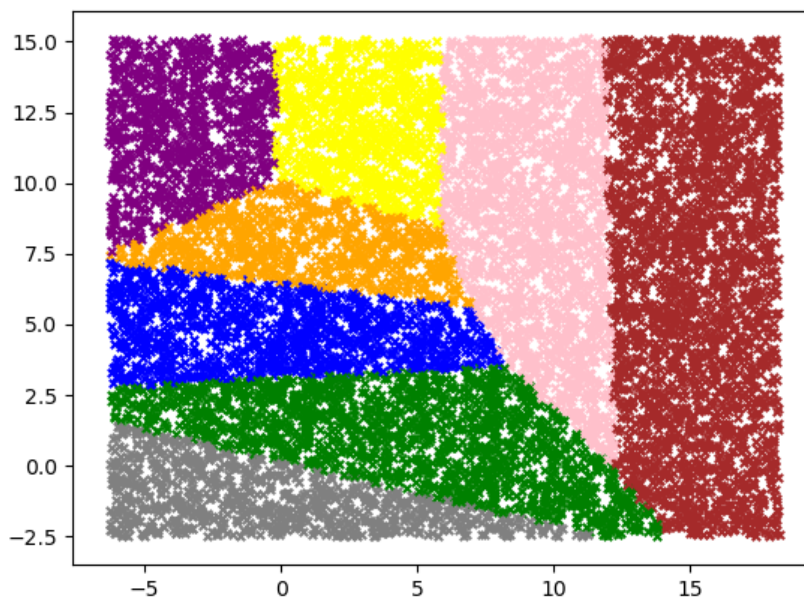
**:2clstrain1200**



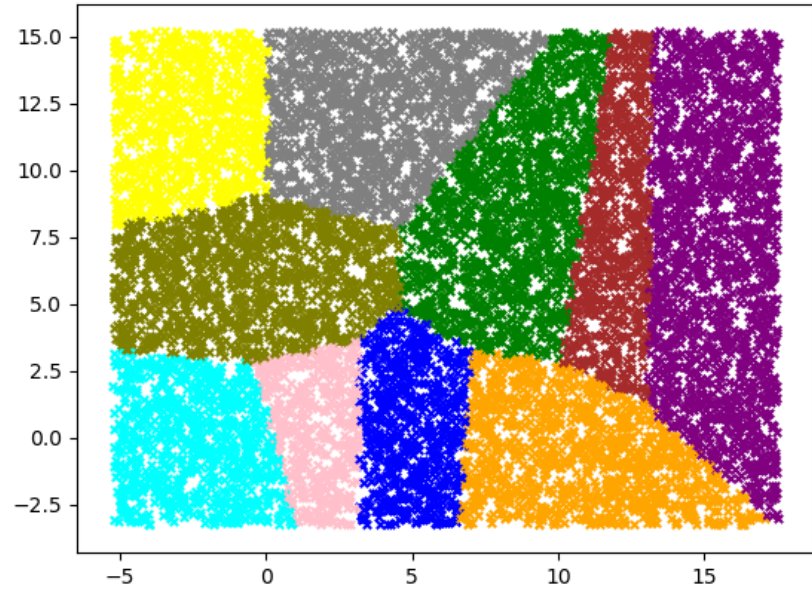
$r = 0.1$   
 $m = 10$   
دقت داده های آموزشی: 0.9761  
دقت داده های تست: 0.9722



$r = 1$   
 $m = 8$   
دقت داده های آموزشی: 0.9714  
دقت داده های تست: 0.9638



$r = 0.1$   
 $m = 10$   
دقت داده های آموزشی: 0.9504  
دقت داده های تست: 0.9466



$r = 1$   
 $m = 6$   
دقت داده های آموزشی: 0.9533  
دقت داده های تست: 0.9511

