



دانشگاه صنعتی امیرکبیر
(پلی تکنیک تهران)
دانشکده ریاضی و علوم کامپیوتر

سری چهارم تمرین تدریسیاری درس داده کاوی
خوشه بندی، شبکه عصبی، کاوش الگوهای پرتکرار

استاد درس: دکتر شاکری

طراح: امید سقط چیان

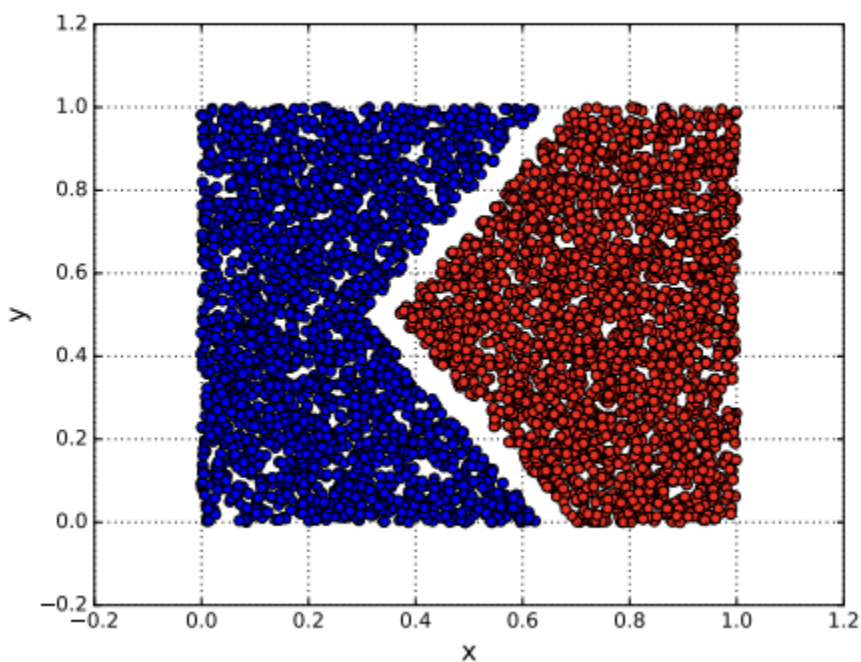
تاریخ انتشار: ۵ خرداد ۱۴۰۲

مهلت تحویل: قبل از امتحان پایانترم

سوال یک:

شبکه عصبی رسم کنید که دو گروه از داده های زیر را بتواند به درستی دسته بندی کند.

- چه تعداد مرز خطی نیاز داریم؟
- از کمترین تعداد لایه های مخفی و نورون ها استفاده کنید.
- تمامی پارامترها را مشخص کنید. (لازم نیست که محاسبات دقیق باشند اما باید مشخص کنید که هر کدام از نورون ها کدام ناحیه را مشخص میکنند و پارامترها به صورت تقریبی درست باشند).
- بگویید از کدام تابع فعالساز استفاده میکنید.



سوال دو:

الگوریتم k-means را با معیار فاصله اقلیدسی روی داده های زیر پیاده سازی کنید.

$A1=(2,10)$, $A2=(2,5)$, $A3=(8,4)$, $A4=(5,8)$, $A5=(7,5)$, $A6=(6,4)$, $A7=(1,2)$, $A8=(4,9)$

برای سادگی کار، فواصل در جدول زیر آمده است.

	A1	A2	A3	A4	A5	A6	A7	A8
A1	0	$\sqrt{25}$	$\sqrt{36}$	$\sqrt{13}$	$\sqrt{50}$	$\sqrt{52}$	$\sqrt{65}$	$\sqrt{5}$
A2		0	$\sqrt{37}$	$\sqrt{18}$	$\sqrt{25}$	$\sqrt{17}$	$\sqrt{10}$	$\sqrt{20}$
A3			0	$\sqrt{25}$	$\sqrt{2}$	$\sqrt{2}$	$\sqrt{53}$	$\sqrt{41}$
A4				0	$\sqrt{13}$	$\sqrt{17}$	$\sqrt{52}$	$\sqrt{2}$
A5					0	$\sqrt{2}$	$\sqrt{45}$	$\sqrt{25}$
A6						0	$\sqrt{29}$	$\sqrt{29}$
A7							0	$\sqrt{58}$
A8								0

$K = 3$ در نظر بگیرید. همچنین centroid های $A1$, $A4$, $A7$ خواهند بود. الگوریتم را تنها یک مرحله انجام دهید. بعد

از این مرحله نشان دهید:

- هر کدام از اعضا متعلق به کدام خوشه هستند
- مرکز هر خوشه را تعیین کنید
- در نمودار نقاط و خوشه ها را نشان دهید.

سوال سه:

همان داده های سوال قبل را در نظر بگیرید. اینبار میخواهیم با الگوریتم DBScan داده ها را خوشه بندی کنیم. فرض کنید که اپسیلون و minpoints برابر است با ۲.

- نقاط هسته، مرزی و نویز را مشخص کنید.
- خوشه ها را تعیین کنید.
- همانند سوال قبل روی نمودار نشان دهید که خوشه بندی ها به چه صورت خواهند بود.
- اگر اپسیلون برابر با $\sqrt{10}$ باشد. چه تغییری در خوشه بندی ایجاد میشوند. خوشه بندی را رسم کنید و بگویید به چه علت این تغییرات ایجاد شد.

سوال چهار:

الگوریتم Apriori را روی داده های زیر پیاده سازی کنید.

(support threshold = 33% & confidence threshold = 60%)

Transaction ID	Items
T1	{HotDogs, Buns, Ketchup}
T2	{HotDogs, Buns }
T3	{HotDogs, Coke, Chips }
T4	{Chips, Coke }
T5	{Chips, Ketchup }
T6	{HotDogs, Coke, Chips}

- کاندیدهای پرتکرار از سایز ۱ تا ۴ را مشخص کنید.

- قوانین انجمنی استخراج شده را مشخص کنید.
- قوانین انجمنی را با معیار اطمینان ارزیابی کنید.