

## به نام ایزد منان

تمرین دوم درس مبانی هوش محاسباتی، «مبانی سیستم‌های فازی»



استاد درس: دکتر عبادزاده



بهار ۹۹ - دانشکده مهندسی کامپیوتر، دانشگاه صنعتی امیرکبیر

نکاتی در مورد این تمرین نیاز به توجه و دقت دوستان دارد.

۱- در جدول زیر نحوه اعمال جریمه تاخیر در ارسال تمرین‌ها ذکر شده است.

میزان جریمه	میزان تاخیر (روز)
هر روز ۵٪	۱ الی ۲ روز
هر روز ۱۰٪	۲ الی ۶ روز

در صورتی که بین ۷ تا ۱۴ روز تاخیر داشته باشید، نمره شما از ۵۰٪ محاسبه می‌شود و پس از این بازه با توجه به سایر تمرین‌ها و زمان تحویل، به تمرین ارسالی نمره‌ای تعلق نمی‌گیرد.

۲- هرگونه کپی کردن باعث عدم تعلق نمره به تمامی افراد مشارکت کننده در آن می‌شود.

۳- آخرین مهلت ارسال تمرین، ساعت ۲۳:۵۵ دقیقه روز **دوشنبه ۲۶ خرداد** می‌باشد. این زمان با توجه به جمع‌بندی‌های صورت گرفته، شرایط و با توجه به محدودیت زمانی در نظر گرفته شده است و **قابل تمدید نمی‌باشد**.

۴- دوستان فایل ارسالی خود را به صورت فشرده و به صورت «شماره دانشجویی\_HW03» مانند HW02\_97131123 نام گذاری کنید. در این فایل باید مواردی نظیر کدها، فایل پی‌دی‌اف گزارش و سایر موارد مورد نیاز در هنگام بررسی وجود داشته باشد و صرفاً این فایل در روز ارائه در نظر گرفته می‌شود.

۵- این تمرین دارای تحویل حضوری می‌باشد. زمان آن متعاقباً در مودل درس اعلام می‌شود.

۶- کدهای خود را به صورت مناسب کامنت گذاری کنید. به صورتی که بتوان حداقل روال اجرا و موارد مورد نیاز را درک کرد.

۷- سعی کنید ابتدا تمامی سوالات و بخش‌ها را مطالعه کنید.

۸- در این تمرین بخش‌هایی **امتیازی و اختیاری** در نظر گرفته شده است. این بخش‌ها با علامت ✨ مشخص شده‌اند.

۹- در صورت هرگونه سوال یا مشکل می‌توانید با تدریس‌یاران درس از طریق ایمیل زیر در ارتباط باشید.

ci.spring2020@gmail.com

بخش اول - مباحث تئوری و مسائل تشریحی (۶۰ نمره + ۵ نمره اضافی)

۱- با ذکر مثال، توضیح دهید که متغیر فازی چه تفاوتی با متغیر تصادفی دارد؟

۲- مجموعه های فازی زیر را در نظر بگیرید:

$$A = \left\{ \frac{0.1}{2}, \frac{1}{5}, \frac{0.7}{3} \right\}, \quad B = \left\{ \frac{0.5}{2}, \frac{0.3}{6} \right\}, \quad C = \left\{ \frac{0.2}{6}, \frac{0.8}{2}, \frac{1}{4} \right\}$$

$$X = \{2, 3, 4, 5, 6\}$$

$$Y = \{2, 3, 6\}$$

که  $X$  مجموعه مرجع  $(A \text{ و } C)$  و  $Y$  مجموعه مرجع  $(B)$  است. مطلوب است محاسبه موارد زیر با ذکر کامل مراحل:

(الف)  $(B \cup C) \cap A$

(ب)  $A \cap \bar{A}$

(ج)  $A \cap B$

(د)  $A \times B$

۳- همانطور که می دانید مجموعه عملگرهایی را کلاس نرمال می نامیم که خاصیت جابه جایی، شرکت پذیری، توزیع پذیری و دموگن داشته باشند. حال اگر در مجموعه های فازی، عملگرهای اشتراک، اجتماع و متمم را مطابق آنچه در زیر آمده است در نظر بگیریم، با بررسی وجود هر کدام از خواص مطرح شده، بیان کنید که آیا می توان این عملگرها را یک کلاس نرمال در نظر گرفت؟

$$\mu_{A \cap B}(x) = \mu_A(x) * \mu_B(x)$$

$$\mu_{A \cup B}(x) = \max(\mu_A(x), \mu_B(x))$$

$$\mu_{\bar{A}}(x) = 1 - \mu_A(x)$$

۴- دو رابطه  $R$  و  $S$  را در نظر بگیرید و به سوالات زیر پاسخ دهید.

(الف) با ذکر مراحل، جداپذیر بودن یا نبودن هر کدام از رابطه ها را مشخص نمایید.

(ب) رابطه  $Z = R \circ S$  را محاسبه نمایید. (ترکیب بر اساس  $\max\text{-min}$ )

(ج) با داشتن رابطه  $Z$  در قسمت ب و ورودی  $A_1 = \left\{ \frac{0.1}{a_1}, \frac{0.8}{a_3} \right\}$ ، خروجی را به دست آورید. (ترکیب بر اساس  $\max\text{-product}$ )

<b>R</b>	$b_1$	$b_2$	$b_3$
$a_1$	۰.۳	۰.۷	۰.۷
$a_2$	۰.۳	۰.۸	۰.۹
$a_3$	۰.۳	۰.۸	۱

<b>S</b>	$c_1$	$c_2$	$c_3$
$b_1$	۰.۲	۰.۷	۰.۵
$b_2$	۰.۳	۰.۸	۰.۹
$b_3$	۰.۴	۰.۱	۰.۴

۵- رابطه  $R = \{(x_1, x_2, y) | x_1, x_2, y \in R, x_1 + x_2 + y^2 = 10\}$  را در نظر بگیرید. با توجه به ورودی های فازی  $x_1 = A_1$  و  $x_2 = A_2$  خروجی فازی  $(y = B)$  را به دست آورید.

$$A_1 = \left\{ \frac{0.5}{1}, \frac{0.3}{0} \right\}, \quad A_2 = \left\{ \frac{0.3}{0}, \frac{0.7}{1} \right\}$$

۶- دو مجموعه‌ی فازی A, B را در نظر بگیرید که هر دو بر روی مجموعه U تعریف شده‌اند:

$$U = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7\}$$

$$A = \left\{ \frac{0.3}{x_1}, \frac{0.5}{x_2}, \frac{1}{x_3}, \frac{0.7}{x_4}, \frac{0.6}{x_5} \right\}$$

$$B = \left\{ \frac{0.2}{x_1}, \frac{0.5}{x_2}, \frac{0.3}{x_3}, \frac{0.8}{x_4} \right\}$$

برش لامبدا از روش‌هایی است که در عملیات غیرفازی‌سازی<sup>۱</sup> برای تبدیل مجموعه‌های فازی به مجموعه‌های crisp کاربرد دارد. برش‌های لامبدای خواسته شده بر روی مجموعه‌ها اعمال کرده و خروجی آن را به دست آورید.

$$C = A' \cap B, \lambda = 0.3$$

$$B, \lambda = 0.5, 0$$

۷- در این سوال می‌خواهیم ویژگی‌هایی از تابع تعلق فازی<sup>۲</sup> را بررسی کنیم.

الف) در مورد ویژگی normal یا subnormal بودن مجموعه‌های فازی و همچنین convexity و nonconvexity مجموعه‌های فازی تحقیق کرده و به صورت خلاصه و با رسم نمودار تابع تعلق‌های فرضی توضیحشان دهید (رسم دو نمودار هم کفایت می‌کند).

ب) با تعریفی که از اشتراک و اجتماع مجموعه‌های فازی داریم، آیا اشتراک دو مجموعه فازی که Convex هستند، مجموعه‌ای فازی خواهد شد؟ اجتماعشان چطور؟ برای ویژگی normal بودن هم این عملگرها را بررسی کنید.

ج) ناحیه‌های Core, Boundary و Support را هم روی نمودارهای قبلی رسم شده یا به روی یک نمودار جدید مشخص کرده و مختصراً توضیح دهید.

۸- در آخرین قسمت از تمرین تئوری فازی که شامل این سوال و سوال بعدیست، هدف درک مفهوم و کاربردی است که قوانین فازی در دنیای واقعی می‌توانند داشته باشند.

در اینجا قصد داریم از ویژگی‌هایی که یک نوع بیماری و پیروسی از خودش به جا می‌گذارد به درجه‌ی بیماری در بدن این فرد پی ببریم. پس از پردازش تصویر بر روی سلول‌های آسیب دیده، دو ویژگی را می‌توانیم استخراج کنیم:

۱- تعداد نقاط سیاه مشاهده شده داخل سلول

۲- شکل این نقاط سیاه مشاهده شده

از آنجایی که مشخص کردن دقیق این دو ویژگی از روی این تصاویر کار دشواری است، ملزم به بیان فازی این متغیرها و مفاهیم هستیم. با استفاده از قوانین و روابط فازی به روی استانداردهای از پیش تعریف شده که قبلاً با دقت و هزینه‌ی بالا توسط افراد خبره به دست آمده‌اند، درجه‌ی بیماری را به دست آوریم. مجموعه‌های فازی تعریف شده در سه حوزه‌ی مختلف به صورت زیر می‌باشد.

تعداد نقاط:

$$\text{Low} = \left\{ \frac{1}{0}, \frac{0.4}{10}, \frac{0.2}{20}, \frac{0.1}{30} \right\}, \text{Medium} = \left\{ \frac{0.7}{10}, \frac{0.9}{20}, \frac{0.4}{30}, \frac{0.15}{40} \right\}, \text{High} = \left\{ \frac{0.3}{20}, \frac{0.7}{30}, \frac{0.9}{40}, \frac{1}{50} \right\}$$

شکل نقاط:

$$\text{Circular} = \left\{ \frac{0.8}{R_1}, \frac{0.6}{R_2}, \frac{0.2}{R_3} \right\}, \text{Oval} = \left\{ \frac{0.1}{R_1}, \frac{0.35}{R_2}, \frac{0.7}{R_3}, \frac{0.9}{R_4} \right\}$$

درجه‌ی بیماری:

$$\text{Low-Grade} = \left\{ \frac{0.9}{1}, \frac{0.6}{2}, \frac{0.3}{3}, \frac{0.2}{4}, \frac{0.2}{5} \right\}, \text{High-Grade} = \left\{ \frac{0.6}{3}, \frac{0.8}{4}, \frac{0.9}{5}, \frac{1}{6}, \frac{1}{7} \right\}$$

<sup>1</sup> Defuzzification

<sup>2</sup> Features of Membership Function

حال خروجی فازی را برای قوانین زیر با اعمال کردن ورودی مشاهده شده ی  $R_2$  (شکل نقاط) و ۲۰ (تعداد نقاط) به دست آورید.

- الف) اگر تعداد نقاط زیاد باشد آنگاه درجه ی بیماری بالا است. (با استفاده از max-min)  
ب) اگر شکل نقاط بیضوی شکل باشد آنگاه درجه ی بیماری پایین است. (با استفاده از max-product)  
ج) اگر تعداد نقاط متوسط و شکل نقاط دایروی باشند آنگاه درجه ی بیماری بالا است. (با استفاده از max-Min)

۹- پیش از این با برش لامبدا که مجموعه های فازی را به مجموعه هایی غیرفازی تبدیل می کرد آشنا شده ایم. حال قصد داریم شیوه ی تبدیل خروجی فازی به تنها یک عدد اسکالر را بررسی کنیم. برای این منظور می خواهیم خروجی فازی تولید شده در یکی از بخش های مسئله ی قبل را، غیر فازی سازی کرده، به طوری که مقداری اسکالری برای Grade بیماری به دست بیاوریم.

الف) ابتدا در مورد روش های Weighted Average, Centroid Method, Max Membership Principle, Mean Max Membership Method تحقیق کرده و به طور خلاصه با بیان ویژگی هایشان آن ها را با یکدیگر مقایسه کنید.

ب) پاسخ فازی که برای یکی از قوانین داده شده به دست آورده اید را انتخاب کرده و خروجی که به صورت گسسته به دست آورده اید را در این قسمت، پیوسته فرض کنید. برای این منظور یک توزیع دلخواه که منطبق با مجموعه فازی به دست آمده تان است، رسم کنید. **(راهنمایی: برای ساده شدن کارتان می توانید خروجی نمودارتان را به صورت مثلثی، دوزنقه ای یا ترکیبی از آن ها بکشید.)** سپس خروجی عددی به ازاء هر کدام از روش های ذکر شده به دست آورید. حتی اگر خروجی که انتخاب کرده اید بنظرتان توزیع پیچیده ای دارد، مقادیرش را با اعداد نزدیک تقریب زده تا بتوانید از روش های ذکر شده به راحتی استفاده کنید. زیرا هدف این قسمت صرفا استفاده و آشنایی شما با روش های غیرفازی سازی است.

در این بخش علاوه بر کد و نمودارها، پاسخ تشریحی و تحلیل شما حائز اهمیت است.

در این قسمت، قصد داریم خوشه بندی با الگوریتم FCM<sup>۳</sup> را پیاده سازی نماییم که روابط مورد نیاز را در اینجا<sup>۴</sup> می‌توانید مطالعه کنید. داده‌گان‌های مورد نیاز نیز در کنار همین فایل در مدل در پوشه‌ای به نام data set بارگذاری شده است.

در گزارش خود ابتدا به موارد زیر پاسخ دهید:

۱. شرح مختصری در مورد مراحل و نحوه کار کلی الگوریتم و برخی موارد جزئی تر مانند نوع رویکردتان در مقداردهی اولیه ماتریس تعلق‌ها و تعیین شرط خاتمه الگوریتم ارائه دهید.

۲. رابطه تابع هزینه<sup>۵</sup> در خوشه خوشه‌بندی FCM را بنویسید و با توجه به رابطه توضیح دهید که مقدار هزینه به دست آمده از لحاظ مفهومی چه چیزی را نشان می‌دهد؟ مقدار هزینه با افزایش تعداد خوشه‌ها چه تغییری می‌کند؟ چرا؟

۳. پارامتر m (در رابطه مراکز، مقدار تعلق‌ها و تابع هزینه) چه عددی در نظر گرفته شده است؟

۴. تعداد خوشه‌ای که منجر به یک خوشه بندی خوب شود، تعداد خوشه بهینه است. در مورد معیارهای یک خوشه بندی فازی خوب و نحوه ارزیابی<sup>۶</sup> آن تحقیق کنید و توضیح دهید.

حال پس از اجرای الگوریتم FCM بر روی هر داده‌گان، موارد زیر را گزارش کنید:

- **تعداد خوشه‌ها و رویکردتان در تعیین آن به عنوان پارامتر مسئله:** این بخش از اهمیت زیادی برخوردار است و نوع نگاه و خلاقیت شما در این بخش، تاثیر مستقیمی در نمره شما دارد. توجه داشته باشید که ممکن است رویکردتان برای تعیین تعداد خوشه بهینه برای داده‌گان‌های مختلف، با هم متفاوت باشد.

اگر ایده‌ای برای این کار نداشتید، به عنوان یک رویکرد حریصانه می‌توانید از روش elbow<sup>۷</sup> استفاده کنید. (به ازاء تعداد خوشه‌های مختلف، مقدار تابع هزینه متناظر با آن را حساب کنید و نمودار تعدادخوشه – هزینه را رسم کنید. بهترین تعداد خوشه جایی است که نمودار شکستگی پیدا می‌کند!)

- **به عنوان یک مورد امتیازی، در مورد یافتن بهترین تعداد خوشه در خوشه بندی FCM (و نه k-means) تحقیق کنید و با ذکر منبع و توضیح مختصر ایده از روش ارائه شده استفاده کنید.** (مثلا به عنوان پیشنهاد نگاهی به این مقاله<sup>۸</sup> یا این مقاله<sup>۹</sup> بیندازید)

- **مقدار هزینه (توسط تابع هزینه)**

در مورد داده‌گان با داده‌های دو بعدی، علاوه بر گزارش موارد بالا، ترسیم‌های زیر مورد نیاز است:

- **پلات کردن کل نقطه‌ها + پلات کردن مراکز خوشه‌های به دست آمده؛ متفاوت با سایر نقطه‌ها که مشخص باشند!**
- **نمایش مرزبندی خوشه‌ها :**

بعد از اتمام الگوریتم FCM و به دست آمدن مراکز، برای رسم مرزهای خوشه‌ها ساده ترین کاری که می‌توان انجام داد به صورت زیر است:

فرض کنید که داده‌ها در راستای مولفه‌ی اول (محور x) بین ۲.۵ تا ۳.۷۵ پخش شده‌اند و در راستای مولفه‌ی دوم (محور y) بین ۱ تا ۵.۵؛ برای دیدن مرزهای خوشه‌ها کافیست که تعداد زیادی داده به صورت تصادفی یکنواخت،

<sup>3</sup> Fuzzy C-Means (FCM)

<sup>4</sup> [https://home.deib.polimi.it/matteucc/Clustering/tutorial\\_html/cmeans.html](https://home.deib.polimi.it/matteucc/Clustering/tutorial_html/cmeans.html)

<sup>5</sup> Cost function

<sup>6</sup> Evaluation

<sup>7</sup> [https://en.wikipedia.org/wiki/Elbow\\_method\\_\(clustering\)](https://en.wikipedia.org/wiki/Elbow_method_(clustering))

<sup>8</sup> <https://www.sciencedirect.com/science/article/abs/pii/S0031320310002013>

<sup>9</sup> <https://www.sciencedirect.com/science/article/abs/pii/S0167865504003629?via%3Dihub>

بین این بازه‌ها تولید کنید (به طوری که تقریباً صفحه را بپوشاند) و سپس این داده‌ها را به FCM بدهید. برای هر داده یک خروجی فازی داده می‌شود. (با توجه به اینکه مراکز از قبل به دست آمده‌اند، برای یافتن تعلق داده جدید به هر کدام از خوشه‌ها از همان رابطه Uij استفاده کنید.) خوشه‌ی هر داده را خوشه‌ای که داده بیش‌ترین تعلق به آن را دارد در نظر بگیرید و هر خوشه را با یک رنگ مجزا معین کنید (داده‌های متعلق به آن خوشه را با رنگ مخصوص آن خوشه پلات کنید). با این رویکرد و از آنجایی که این تعداد زیاد داده‌ها عملاً کل فضا را می‌پوشانند، مرزبندی خوشه‌ها مشخص می‌شود.

**موفق باشید**