

تکالیف به صورت پی دی اف با عنوان HW1 به ایمیل [rikhtehgaran\\_iut@yahoo.com](mailto:rikhtehgaran_iut@yahoo.com) ارسال شود. کدهای مربوط نیز به صورت بخش بندی شده و مشخص در قالب Notepad، جداگانه ارسال شود. لازم است در صفحه اول اسامی اعضای گروه سه نفره اعلام شود. توجه: نهایت نمره اضافه بر ۲۰ از این تکلیف ۱/۵ نمره از طریق سوالات مشخص شده و به صورت انفرادی و تحت شرایط عنوان شده در سوالات است!

- ۱- تابعی بنویسید که اطلاعات مجموعه داده house را بگیرد و بر اساس مشخصاتی که کاربر در نظر دارد لیستی از خانه ها را به ترتیب متغیر قیمت چاپ نماید.
  - تابع باید این قابلیت را داشته باشد که برای هر تعداد ویژگی موجود در مجموعه داده که مدنظر کاربر هست و به هر نحو که کاربر مایل بود مانند تعیین مقدار دقیق ویژگی، تعیین بازه برای ویژگی مورد نظر و یا حد پایین و یا بالا و یا حتی انتخاب نکردن هر ویژگی قابل اجرا باشد.
  - **بخش امتیازی سوال:** چنانچه بتوانید داشبوردهای گرافیکی برای عملکرد و نمایش تابع مورد نظر بنویسید (۰/۵ تا ۱ نمره اضافه بر ۲۰ برای آن در نظر گرفته می شود (نمره انفرادی است و به شرطی اختصاص داده می شود که کد نوشته شده توسط شما شباهتی با کد سایر افراد نداشته باشد!))

- ۲- فرض کنید که میانگین وزنی  $\bar{y}_w = \sum_{i=1}^n w_i y_i$  مدنظر باشد که در آن  $w_i$  ها معرف وزن ها هستند و به صورت مقادیر مشخصی بین صفر و یک هستند به طوری که مجموع وزن ها برابر یک هست و  $y_i$  ها از توزیع برنولی پیروی می کنند. با رسم هیستوگرام مشاهدات میانگین وزنی، تحقیق کنید که آیا توزیع نرمال برای میانگین وزنی متغیرهای دودویی تقریب بهتری است یا توزیع پتا؟ پاسخ خود را با اندازه های نمونه  $n$  مختلف بررسی کنید.
  - برای این منظور، مطالعه شبیه سازی ای را طراحی کنید. منحنی توزیع های نرمال و بتا را نیز بر هیستوگرام مشاهدات تولید شده رسم کنید.

- ۳- خوشه بندی K-میانگین را در نظر بگیرید.
  - الف- برنامه را برای حالتی بنویسید که بتواند واحد های آزمایشی که هر یک بردار دو متغیره از ویژگی ها دارند را خوشه بندی کند.
  - ب- چگونه کیفیت عملکرد K-میانگین را با استفاده از داده های شبیه سازی شده می سنجید؟
  - ج- **بخش امتیازی سوال:** الگوریتم را طوری بنویسید که  $n$  مشاهده تک متغیره را بتواند به  $n/a$  خوشه هر یک با  $a$  عضوی که بیشترین شباهت به یکدیگر دارند، خوشه بندی کند. (۰/۵ تا ۱ نمره بسته به اینکه کد ارائه شده شبیه به کد فرد دیگری نباشد)

- ۴- **سوال امتیازی (۰/۵ تا ۱ نمره):** سیستم های توصیه گر (Recommendation systems) کد مربوط به یک سیستم توصیه گر ایتم محور را بنویسید. کد نوشته شده طوری عمل کند که میزان اهمیت هر ایتم که توسط کارفرما مشخص شده هست را نیز به نحو مناسبی در مدلسازی وارد کند!
  - در صورتی نمره اضافه تخصیص می یابد که کد شما شبیه به کد فرد دیگری نباشد و همچنین شباهتی به کدهای آماده ای که برای این توصیه گر ها وجود دارد نیز نباشد.
  - لازم است مجموعه داده مناسبی را پیدا کنید و یا آن را شبیه سازی کنید.
  - آیا می توانید آن را به صورت داشبورد ارائه دهید؟! (۰/۵ تا ۱ نمره)