

یا علی یا عظیم/ به نام آفریننده بهار

تمرین دوم

دانشکده مهندسی صنایع و سیستم‌های مدیریت، دانشگاه صنعتی امیرکبیر

درس داده‌کاوی: مدل‌ها، الگوریتم‌ها و کاربردها، موعد تحویل: ۲۵ فروردین ۱۴۰۳

---

**مساله یک.** مجموعه داده pima-indian-diabetes از مخزن داده‌های استاندارد دانشگاه کالیفرنیا (UCI Machine Learning Repository) مورد نظر است. سوالات زیر را پاسخ دهید.

الف- بعد از پیش‌پردازش داده‌ها، با استفاده از روش «جستجوی روبه‌جلو»، بهترین ترکیب ویژگی‌ها را بدست آورید. از روش «نزدیکترین همسایگی» جهت دسته‌بندی استفاده کنید. بدین‌منظور از روش holdout استفاده کنید. مقدار معیارهای دقت، حساسیت و اختصاصیت را گزارش کنید.

ب- حال از روش kNN جهت دسته‌بندی استفاده کنید. ۶۵ درصد داده‌ها را برای آموزش و الباقی را برای تست به‌کار ببرید. ابتدا کای مناسب را بر اساس دقت بدست آورید. سپس ماتریس در هم ریختگی را تشکیل داده و نتایج مندرج در آن را بر اساس معیارهای دقت، حساسیت (sensitivity) و اختصاصیت (specificity) شرح دهید. همچنین در ادامه با روش leave-one-out عملکرد روش kNN را ارزیابی نموده و معیارهای دقت، حساسیت و اختصاصیت را گزارش کنید.

ج- با استفاده از دسته‌بندی بیزی داده‌ها را دسته‌بندی کنید. در صورت نیاز ۶۵ درصد داده‌ها را برای آموزش و الباقی را برای تست به‌کار ببرید.

**مساله دوم:** داده‌های زیر را در نظر بگیرید:

(0,0), (1,0), (0,1), (1,1), (3,3), (4,3), (3,4), (4,4)

داده‌های فوق را در محور مختصات دو بعدی ترسیم نمایید. به صورت چشمی آنها را به دو کلاس تخصیص دهید (برچسب داده‌ها را معین کنید).

ابتدا با کمک روش نزدیکترین همسایگی، مرز دو کلاس را بیابید. سپس با روش پرسپترون مرز دو کلاس را بیابید.

موفق باشید.