

هوالعلیم
تمرین سوم
دانشکده مهندسی صنایع، دانشگاه صنعتی امیرکبیر
درس: داده کاوی

مجموعه داده breast cancer (ویسکانسین) از مخزن داده‌های استاندارد دانشگاه کالیفرنیا (UCI Machine Learning Repository) مورد نظر است. برای هر کدام از مجموعه داده‌ها، سوالات زیر را پاسخ دهید.

الف- ابتدا داده‌ها را با روش‌های سلسله مراتبی ادغامی و خوشه بندی کامیانگین در دو خوشه قرار دهید. عملکرد روشها را با معیارهای مختلف ارزیابی کنید.

ب- داده‌ها را با روش‌های «شبکه عصبی پرسپترون چند لایه» دسته بندی کنید. روش خود برای یافتن ساختار مناسب شبکه را بیان کنید. از روش اعتبارسنجی سه بخشی استفاده کنید و نتایج را گزارش کنید.

ج- حال داده‌ها را با «شبکه عصبی با تابع محرک شعاعی» دسته بندی کنید. پارامترهای توابع محرک شعاعی (تعداد نرون‌ها، مراکز و پهنای توابع شعاعی) را با روش خوشه بندی کامیانگین بدست آورید. ۶۰ درصد داده‌ها را برای آموزش و الباقی را برای آزمایش به کار ببرید. ماتریس در هم ریختگی را تشکیل داده و نتایج مندرج در آن را بر اساس شاخص‌های مختلف شرح دهید.