



تمرین درخت تصمیم

داده مورد استفاده:

برای انجام این تمرین می‌توانید از مجموعه داده adult از دیتا بیس [UCI](#) استفاده کنید. این مجموعه داده دارای 14 ویژگی است که 6 ویژگی پیوسته و بقیه گسسته می‌باشند. برای سهولت کار می‌توانید از داده‌هایی که فاقد مقادیر برخی ویژگی‌ها هستند صرف نظر کنید. هدف یادگیری این است که بر اساس سایر ویژگی‌ها پیش‌بینی شود که آیا یک شخص دارای درآمدی بیش از 50k است یا خیر.

1. مجموعه مثالهای adult را بصورت تصادفی به 10 مجموعه مساوی ولی غیر تکراری تقسیم کنید. به ازای هر یک از 9 مجموعه حاصل یک درخت تصمیم ایجاد کرده و از مجموعه دهم برای تست استفاده کنید. نموداری از خطاهای حاصل را رسم کرده و روی نتیجه بحث کنید.

2. عملکرد درخت را بر اساس تعداد مثالهای آموزشی بصورت گراف رسم کنید. (منحنی آموزش)

3. نتیجه را برای درخت‌های هرس شده و هرس نشده مقایسه کنید. اختلاف بین تعداد ویژگی‌های درخت هرس شده و نشده را مشخص کنید.

4. برای انتخاب ویژگی مناسب برای هر گره علاوه بر روش استفاده از Information Gain روشهای زیر را نیز پیاده‌سازی کنید:

- تصادفی Random که در آن یکی از ویژگی‌ها بصورت تصادفی انتخاب می‌شوند.
- نسبت بهره (Gain Ratio)

-
- فایل‌های کد و پاسخها را به صورت یک فایل فشرده بر روی ایلرن آپلود کنید.
 - پاسخ‌ها به صورت تایپ شده ارسال شود و پاسخ‌ها به زبان فارسی باشد.
 - به پاسخ‌های مشابه نمره‌ای داده نمی‌شود و صفر برای آنها لحاظ خواهد شد.
 - فرمت فایل ارسالی به صورت HW02.FistName.LastName_StudentID باشد.
 - این تمرین بعد از مهلت تعیین شده تحویل گرفته نمی‌شود.