

یادگیری ماشین - نیمسال اول ۹۵-۱۳۹۴
تکلیف شماره ۱ - تحویل سه شنبه ۱۳۹۴/۷/۲۸

آشنایی با درختهای تصمیم گیری، و یادگیری مفهوم

۱- نرم افزار [Weka](#) که نرم افزاری است که الگوریتمهای متنوع یادگیری در آن پیاده سازی شده است را نصب نموده و با قابلیت‌های مختلف آن آشنا شوید. در این الگوریتم C4.5 بنام J48 وجود دارد.

۲- داده های جدول ۱ وضعیت گذراندن درس یادگیری ماشین توسط تعدادی از دانشجویان نشان می دهد. وضعیت هر دانشجو توسط ۸ خصیصه توصیف می شود. این خصائص و مقادیر آنها عبارتند از:

- حضور فعال در کلاس: خوب، متوسط، ضعیف
- مطالعه هفتگی کتاب: خوب، متوسط، ضعیف
- مطالعه از روی جزوه: خوب، متوسط، ضعیف
- میان نیمسال: خوب، متوسط، زیر ۵۰
- پایان نیمسال: خوب، متوسط، زیر ۵۰
- تکالیف: خوب، متوسط، ضعیف
- تحقیق: خوب، متوسط، ضعیف
- پروژه: خوب، متوسط، ضعیف

خصیصه هدف نمره نهائی با مقادیر خوب، متوسط، افتاده می باشد. اعمال زیر را انجام دهید:

- داده های جدول ۱ را به فرمت مناسب برای ورود به نرم افزار وکا (arff) آماده نمائید.
- بدون هرس کردن درخت تصمیم گیری مرتبط با این داده های آموزشی را یاد بگیرید.
- درخت تصمیم گیری یاد گرفته شده را بر روی دو نمونه آزمون جدول ۲ آزمایش نمائید.
- درخت را با استفاده از قوانین نمایش دهید.
- درخت تصمیم گیری را با استفاده از هرس کردن بیاموزید و نتیجه آن را روی داده های آزمون بررسی کنید.
- خودتان با تغییراتی که لازم می داند اثر بیش پوشش را بررسی کرده و نشان دهید که چرا بیش پوشش انجام گرفته است و چگونه می توان جلوی آن را گرفت.
- به دلخواه خود یکی از خصیصه ها را تبدیل به خصیصه پیوسته نمائید و دوباره درخت تصمیم گیری را یاد بگیرید.
- اثر داده های خطا دار و یا خصیصه های بدون مقدار را بررسی و گزارش کنید.

جدول ۱ داده های آموزشی

فرد	حضور فعال در کلاس	مطالعه هفتگی کتاب	مطالعه از روی جزوه	میان نیمسال	پایان نیمسال	تکالیف	تحقیق	پروژه	نمره نهائی
۱	خوب	خوب	خوب	خوب	خوب	خوب	خوب	خوب	خوب
۲	خوب	خوب	خوب	خوب	خوب	ضعیف	ضعیف	ضعیف	متوسط
۳	ضعیف	ضعیف	خوب	زیر ۵۰	خوب	خوب	خوب	خوب	افتاده
۴	خوب	ضعیف	متوسط	متوسط	متوسط	ضعیف	متوسط	خوب	متوسط
۵	ضعیف	ضعیف	ضعیف	زیر ۵۰	زیر ۵۰	خوب	خوب	خوب	افتاده
۶	متوسط	خوب	ضعیف	خوب	خوب	خوب	خوب	خوب	خوب
۷	خوب	ضعیف	خوب	متوسط	متوسط	خوب	خوب	خوب	متوسط
۸	خوب	خوب	ضعیف	خوب	خوب	متوسط	متوسط	متوسط	خوب
۹	ضعیف	ضعیف	خوب	خوب	زیر ۵۰	خوب	خوب	خوب	افتاده
۱۰	متوسط	خوب	خوب	خوب	متوسط	خوب	ضعیف	خوب	متوسط
۱۱	خوب	ضعیف	خوب	زیر ۵۰	خوب	متوسط	متوسط	متوسط	افتاده
۱۲	ضعیف	متوسط	متوسط	متوسط	متوسط	متوسط	ضعیف	ضعیف	متوسط
۱۳	متوسط	متوسط	خوب	خوب	زیر ۵۰	متوسط	متوسط	متوسط	افتاده
۱۴	خوب	متوسط	خوب	متوسط	متوسط	ضعیف	متوسط	متوسط	متوسط
۱۵	ضعیف	ضعیف	ضعیف	زیر ۵۰	زیر ۵۰	متوسط	متوسط	متوسط	افتاده
۱۶	متوسط	خوب	خوب	خوب	خوب	خوب	خوب	خوب	خوب
۱۷	متوسط	خوب	متوسط	خوب	خوب	خوب	خوب	خوب	خوب
۱۸	متوسط	متوسط	خوب	متوسط	متوسط	متوسط	متوسط	ضعیف	متوسط
۱۹	متوسط	ضعیف	متوسط	زیر ۵۰	متوسط	خوب	متوسط	ضعیف	افتاده
۲۰	خوب	خوب	متوسط	خوب	خوب	خوب	خوب	خوب	خوب
۲۱	متوسط	خوب	خوب	خوب	خوب	خوب	خوب	خوب	خوب
۲۲	خوب	ضعیف	ضعیف	متوسط	متوسط	خوب	متوسط	متوسط	متوسط
۲۳	متوسط	متوسط	خوب	متوسط	زیر ۵۰	خوب	متوسط	ضعیف	افتاده
۲۴	متوسط	خوب	ضعیف	خوب	خوب	خوب	خوب	خوب	خوب
۲۵	متوسط	متوسط	خوب	متوسط	متوسط	ضعیف	متوسط	ضعیف	متوسط

جدول ۲ داده های آزمون

۱	خوب	خوب	خوب	خوب	زیر_۵۰	خوب	خوب	خوب	؟
۲	متوسط	متوسط	خوب	متوسط	خوب	خوب	متوسط	خوب	؟

۲- الف) چه تغییراتی لازم است که در داده های آموزشی داده شود تا بتوان فرضیه ای عطفی بوسیله الگوریتمهای FIND_S یا حذف نامزد یادگرفت؟

ب) پس از انجام تغییرات داده شده، الگوریتم FIND_S چه فرضیه ای را یاد می گیرد؟

ج) این فرضیه داده های آزمون را چگونه دسته بندی می نماید؟

پاسخها به سؤالات فوق و نتایج آزمایشهای خود را در گزارشی به فرمت PDF تحویل نمائید.

لطفاً به نکات زیر توجه نمائید:

- تکالیف تا دو روز بعد از موعد مقرر قابل تحویل می باشند و به ازاء هر روز تأخیر ۱۰٪ از نمره آن کسر می گردد.
- فقط به تکالیفی که در سامانه تحویل داده شوند نمره داده خواهد شد.
- به تکالیف مشابه نمره ای تعلق نخواهد گرفت.

موفق باشید

پالهنک