



آزمون میان‌ترم

نکات مهم آزمون میان‌ترم

- مهلت پاسخ به سوالات این آزمون ۱۲۰ دقیقه است.
- برای این آزمون، ملزم به ارائه گزارش متنی (در قالب PDF) شامل نتایج، توضیحات، تحلیل‌ها و مقایسه‌های کامل هر قسمت هستید.
- در پایان آزمون، لازم است که فایل IPYNB. کدها که در آن نتایج خروجی سلول‌ها قابل مشاهده باشند را به همراه فایل PDF. گزارش خود بارگذاری کنید. هم‌چنین فایل IPYNB. خود را در قالب یک نوتبوک گوگل کولب و از طریق لینک پابلیک آن در صفحه اول گزارش خود، در اختیار قرار دهید
- فایل ارسالی باید یک بسته فشرده در قالب zip. باشد که درون آن، پوشه‌ای با نام «شماره دانشجویی شما» قرار دارد. در این پوشه صرفاً دو فایل —گزارش PDF. و کد IPYNB. — باید موجود باشد؛ هیچ فرمت دیگری تصحیح نخواهد شد.

۱ سوال اول - سوال هماهنگ‌شده (۷.۵ نمره)

به سوالات هماهنگ‌شده‌ای که در این سند قرار داد شده‌اند پاسخ دهید.

۲ سوال دوم (۷.۵ نمره)

درستی یا نادرستی هریک از گزاره‌های زیر را با ذکر دلیل مشخص کنید.

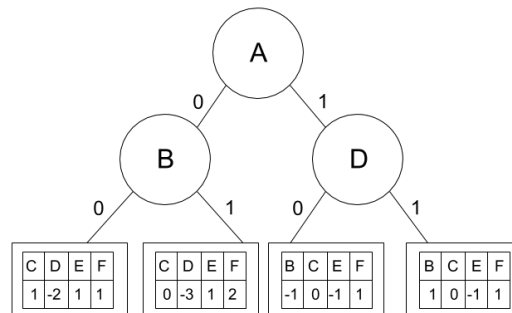
- آ. طبقه‌بند بیز، بهترین طبقه‌بندی است که می‌توان برای جداسازی یک مسئله دوکلاسه طراحی کرد.
- ب. استفاده از رویکرد بیز برای تخمین پارامترهای توزیع، می‌تواند مانع از بیش‌برازش (Overfitting) شود.
- ج. استفاده از معیار Information Gain برای ساخت درخت در شرایطی که بعضی از ویژگی‌ها حالات زیادی دارند، مناسب نیست.
- د. هر شبکه عصبی چندلایه با توابع فعال‌ساز خطی در لایه‌های پنهان می‌تواند به عنوان یک شبکه عصبی بدون هیچ لایه پنهانی نمایش داده شود.
- ه. در برآورد چگالی به‌کمک پارزن با هسته گاوسی، هرچه پهنای پنجره (h) کوچک‌تر انتخاب شود، بایاس تخمین کاهش می‌یابد اما واریانس آن افزایش پیدا می‌کند؛ بنابراین باید بین این دو موازنه برقرار کرد.
- و. در طبقه‌بند پارزن با هسته گاوسی، هنگامی که پهنای پنجره (h) به صفر میل کند، تصمیم‌گیرنده به‌طور نظری به طبقه‌بند «نزدیک‌ترین همسایه» (1-NN) همگرا می‌شود.

۳ سوال سوم (۱۰ نمره)

برای بهره‌برداری از ویژگی‌های مطلوب طبقه‌بندهای درخت تصمیم و پرسپترون، مهدی الگوریتم جدیدی به نام «درخت پرسپترون» ایجاد کرده که ویژگی‌های هر دو را ترکیب می‌کند. درخت‌های پرسپترونی شبیه به درخت‌های تصمیم هستند؛ اما هر leaf node به جای مکانیزم رأی اکثریت، شامل یک پرسپترون است.

برای ایجاد یک درخت پرسپترون؛ اولین مرحله، اجرای یک الگوریتم یادگیری درخت تصمیم معمولی (مانند ID3) و انجام تقسیم‌بندی بر اساس ویژگی‌ها تا رسیدن به عمق حداکثر مشخص شده است. هنگامی که به عمق حداکثر می‌رسیم، در هر leaf node، یک پرسپترون روی ویژگی‌های باقی‌مانده که هنوز در آن شاخه استفاده نشده‌اند، آموزش داده می‌شود. دسته‌بندی یک نمونه جدید از طریق مراحل مشابهی انجام می‌شود. ابتدا نمونه از طریق درخت تصمیم بر اساس مقادیر ویژگی‌هایش گذر می‌کند. وقتی به یک leaf node می‌رسد، پیش‌بینی نهایی با اجرای پرسپترون متناظر در آن گره انجام می‌شود.

فرض کنید که دارای مجموعه داده‌ای با ۶ ویژگی دودویی $\{A, B, C, D, E, F\}$ و دو برجسب خروجی $\{-1, 1\}$ هستید. یک درخت پرسپترون با عمق ۲ روی این مجموعه داده‌ها در شکل ۱ آمده است. وزن‌های پرسپترون نیز در leaf node ها آمده است (فرض کنید که بایاس برای هر پرسپترون $b = 1$ است).



شکل ۱: درخت پرسپترون با عمق دو.

- آ. برای نمونه $x = [1, 1, 0, 1, 0, 1]$ ، درخت پرسپترون داده‌شده چه برجسب خروجی‌ای را پیش‌بینی می‌کند؟
 ب. آیا مرتز تصمیم درخت پرسپترون همواره خطی است؟ برای مقادیر کوچک حداکثر عمق، کیفیت آموزش درخت تصمیم و درخت پرسپترون را با ذکر دلیل مقایسه کنید. آیا تفاوتی دارند؟

۴ سوال چهارم (۳۵ نمره)

مجموعه داده مربوط به طبقه‌بندی پوشش جنگلی در کتابخانه سایکیت‌لرن را فراخوانی کنید (اطلاعات بیشتر).

- آ. کل یا بخشی از داده‌های این مجموعه داده را با نسبتی دلخواه به مجموعه‌های آموزش و آزمون تقسیم کنید.
 ب. یک طبقه‌بند خطی مانند رگرسیون لجستیک را تعریف کرده و روی داده‌ها آموزش دهید. نتایج روی داده‌های آزمون را حداقل با چهار شاخص و ماتریس درهم‌ریختگی نشان داده و تحلیل کنید. در انتخاب فرآپارامترها آزاد هستید؛ اما لازم است که نتایج را به صورت کامل مقایسه و تحلیل کنید. نتایج مربوط به کدام کلاس (ها) بدتر است؟ از کجا فهمیدید و برای حل این مشکل چه پیشنهادی دارید؟ پیاده‌سازی کرده و نتیجه را تحلیل کنید.
 ج. حال برنامه‌ای بنویسید که درخت تصمیمی برای طبقه‌بندی کلاس‌های این مجموعه داده طراحی کند. خروجی درخت تصمیم خود را با برنامه‌نویسی و یا به صورت دستی تحلیل کنید.
 د. با استفاده از ماتریس درهم‌ریختگی و حداقل سه شاخص ارزیابی مربوط به وظیفه طبقه‌بندی، عمل کرد درخت آموزش داده‌شده خود را روی بخش آزمون داده‌ها ارزیابی کنید و نتایج را به صورت دقیق گزارش کنید. تأثیر مقادیر کوچک و بزرگ حداقل دو فرآپارامتر را بررسی کنید. تغییر فرآپارامترهای مربوط به هرس کردن چه تأثیری روی نتایج دارد و مزیت آن چیست؟

۵ سوال پنجم (۴۰ نمره)

یک مجموعه داده شتاب‌سنج برای تشخیص حرکات دست را از طریق [این پیوند](#) در اختیار دارید. برای هر نمونه سه ویژگی acc_x ، acc_y و acc_z اندازه‌گیری شده و هر نمونه به یکی از کلاس‌های زیر تعلق دارد:

۱. حرکت عمودی (بالا/پایین)

۲. حرکت چرخشی

۳. سکون دست

آ. الگوریتم kNN را برای این مجموعه داده پیاده‌سازی کنید. پیاده‌سازی شما باید امکانات زیر را فراهم کند:

- تعریف یک تابع فاصله (به عنوان مثال فاصله اقلیدسی)؛

- امکان تعیین مقدار k توسط کاربر؛

- استفاده از رأی‌گیری اکثریت (majority voting) برای برچسب‌گذاری نمونه‌های آزمون.

سپس معیارهای دقت، (Accuracy) بازیابی (Recall) و صحت (Precision) را برای هر کلاس محاسبه و گزارش کنید. همچنین عملکرد مدل را برای مقادیر مختلف k مقایسه و تحلیل نمایید.

ب. با ثابت نگه داشتن نسبت ۷۰٪ آموزش و ۳۰٪ آزمون، موارد زیر را نیز اجرا کرده و نتایج را گزارش کنید:

(۱) نسخه وزن‌دار الگوریتم kNN؛

(۲) ارزیابی مدل با استفاده از cross-validation.