



تمرین سری ششم

مهلت تحویل: جمعه ۲۱ دی
طراح تمرین: مهدی شفیعی، علی رستمی

نکات ضروری

۱. لازم است کد تمرین به همراه گزارش تحویل داده شود که شامل پاسخ سوالات نظری، توضیحات و تحلیل کد و تصاویر و نمودارهای لازم می باشد.
۲. کد تمرین باید در زبان پایتون، به صورت notebook و شامل نتایج اجرا باشد.
۳. فایل گزارش باید به فرمت pdf باشد. حجم گزارش به هیچ عنوان معیار نمره دهی نیست، لذا لطفا در حد نیاز توضیح دهید.
۴. استفاده از کتابخانه های از پیش آماده مجاز است، مگر اینکه نقیض آن گفته شود. در گزارش کتابخانه ها ذکر شوند.
۵. نمره هر تمرین از ۱۰۰ است (مگر آنکه نمره امتیازی در تمرین مشخص شده باشد). تا سه روز بعد از مهلت اعلام شده برای تمرین امکان بارگذاری وجود دارد اما به ازای هر روز ۱۰ درصد از نمره تمرین کسر خواهد شد.
۶. مشاهده شباهت نامتعارف در گزارش و کد به منزله تقلب می باشد و طرفین مشمول کسر نمره خواهند شد.
۷. برای این تمرین، می توانید گزارش و پاسخ هایتان را به زبان های فارسی و یا انگلیسی تحویل دهید.

مسئله ۱. Ensemble Learning

یک مسأله دسته‌بندی دوکلاسه با برچسب‌های $\{-1, +1\}$ را در نظر بگیرید. مجموعه داده‌های آموزشی

$$\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$$

را در اختیار داریم. می‌خواهیم T دسته‌بند را به صورت زیر با هم ترکیب کنیم:

$$F_T(x) = \sum_{t=1}^T \alpha_t h_t(x)$$

که در رابطه بالا به ازای هر t ، $h_t(x) \in \{-1, +1\}$ و α_t وزن دسته‌بند t -ام است. تابع خطای دسته‌بند F_T به صورت زیر تعریف می‌شود:

$$E = \sum_{i=1}^n (F_T(x_i) - y_i)^2$$

اگر بخواهیم مشابه روش AdaBoost ضرایب α_t را به صورت ترتیبی بدست آوریم، با در نظر گرفتن تابع خطای بالا، مقدار بهینه‌ی α_T را بدست آورید.

[۲۰ نمره]

مسئله ۲. درخت تصمیم

می‌خواهیم از داده‌های زیر استفاده کرده و یک درخت تصمیم برای دسته‌بندی دانشجویان به دو دسته‌ی تنبل (L) و زرنگ (D) بسازیم.

ویژگی ۱	ویژگی ۲	ویژگی ۳	خروجی
N	A	Q	L
N	V	Q	L
N	V	Q	L
U	V	W	L
U	V	W	L
U	A	E	D
N	A	E	D
N	V	E	D
U	A	W	D
U	A	W	D
U	A	W	D

الف. کدام ویژگی به عنوان ریشه درخت انتخاب می‌شود؟ درخت کامل را بکشید.

ب. Accuracy و Recall را برای داده‌های آموزش محاسبه کنید.

ج. با استفاده از کتابخانه‌های موجود پایتون ابتدا دادگانی مشابه داده‌های تمرین تعریف کنید و درخت تصمیم را بسازید.

[۲۰ نمره]

مسئله ۳. مدل مخفی مارکوف

در این تمرین قصد داریم تا الگوریتم‌های مختلف گفته شده در کلاس برای مدل مخفی مارکوف^۱ را بدون استفاده از پکیج‌های آماده پیاده‌سازی کنیم.

در ابتدا این کد پایتون را برای تولید مجموعه داده^۲ دانلود کنید با اجرا کردن این کد ۳ مجموعه داده برای شما تولید می‌شود که به شما لیستی از مشاهدات^۳ و وضعیت اصلی آن روز ارائه می‌دهد که در ادامه شما باید پاسخ‌های خود را برای هر ۳ مجموعه داده آماده کنید و در نهایت مجموعه داده‌های خود را همراه با پاسخ برای این تمرین ارسال کنید.

^۱ Models Markov Hidden
^۲ Dataset
^۳ Observations

در این مسأله حالات ۴ ما آفتابی، ابری و بارانی میباشد که و مشاهدات ما از دما دارای لیبل خیلی زیاد، زیاد، کم و خیلی کم می باشد. برای حل سؤالات به این ۲ جدول نیاز خواهید داشت: ۱. جدول احتمال انتقال حالت ها^۵

بارانی	ابری	آفتابی	/
۰/۱	۰/۳	۰/۶	آفتابی
۰/۲	۰/۴	۰/۴	ابری
۰/۳	۰/۴	۰/۳	بارانی

۲. جدول احتمال مشاهده ها^۶

خیلی کم	کم	زیاد	خیلی زیاد	/
۰/۱	۰/۲	۰/۴	۰/۳	آفتابی
۰/۲	۰/۳	۰/۳	۰/۲	ابری
۰/۱	۰/۲	۰/۴	۰/۳	بارانی

۳. جدول احتمالات اولیه^۷

بارانی	ابری	آفتابی
۰/۲	۰/۳	۰/۵

(آ) با کمک الگوریتم جلورو^۸ و اطلاعات موجود احتمال مشاهده دنباله های موجود درون مجموعه داده های خود را محاسبه کنید.

(ب) با کمک الگوریتم ویتربی^۹ دنباله ی حالات مطلوب برای مشاهدات درون مجموعه داده خود را پیدا کنید.

(ج) در این بخش با کمک الگوریتم باوم-ولش^{۱۰} می خواهیم پارامترهای مدل را پیش بینی کنیم که احتمال انتقال حالت ها و احتمال مشاهدات می باشد.

(د) در این بخش می خواهی به بررسی نتایج به دست آمده در بخش قبل بپردازیم و نتایج بر روی مجموعه داده با ابعاد مختلف را بررسی کنیم همچنین سعی کنید به طور کلی توضیحی درباره نتایج به دست آمده ارائه دهید که هر کدام چه معنایی دارند ارائه نمودار یا سایر موارد شبیه به آن می تواند بسیار مناسب باشند اما در این بررسی دست شما برای ارائه مسائل مختلف کاملاً باز می باشد (یک آنالیز کامل و مناسب می تواند تا ۱۰ نمره اضافه برای شما به همراه داشته باشد).

نکته: در ۳ بخش اول این مسأله نکته اصلی پیاده سازی مناسب کد این الگوریتم ها ملاک اصلی نمره می باشد.

[۶۰ نمره]

[بارم کل: ۱۰۰ نمره]

با آرزوی موفقیت

states^۴
probabilities transition state^۵
probabilities emission^۶
probabilities initial^۷
forward^۸
viterbi^۹
Baum-Welch^{۱۰}