



هوش مصنوعی

بهار ۱۴۰۰

استاد: محمدحسین رهبان

گرددآورندگان: فرشته فرقانی، علی قربانپور

مهلت ارسال: ۲۱ اردیبهشت

HMM و مقدمه‌ای بر یادگیری ماشین

تمرین پنجم بخش دوم

- مهلت ارسال پاسخ تا ساعت ۲۳:۵۹ روز مشخص شده است.
- همکاری و هم‌فکری شما در انجام تمرین مانعی ندارد اما پاسخ‌های هر کس حتما باید توسط خود او نوشته شده باشد.
- در صورت هم‌فکری و یا استفاده از هر منابع خارج درسی، نام هم‌فکران و آدرس منابع مورد استفاده برای حل سوال مورد نظر را ذکر کنید.
- لطفا تصویری واضح از پاسخ‌های سوالات نظری بارگذاری کنید. در غیر این صورت پاسخ شما تصحیح نخواهد شد.

HMM (۵۰ نمره)

۱. (۵۰ نمره) در این مسأله می‌خواهیم از HMM برای دیکود کردن یک دنباله ژنوم استفاده کنیم. همانطور که می‌دانیم یک رشته DNA از یک سری عضو از مجموعه $\{A, C, G, T\}$ تشکیل شده است. حال فرض کنید یک متغیر مخفی مانند S وجود داشته باشد که تولید هر رشته DNA را کنترل کند و دو حالت $\{S_1, S_2\}$ می‌گیرد. حال با فرض احتمالا انتقال و انتشار و آغازین داده شده و با فرض آنکه رشته observe شده برابر با CGTCAG باشد به سوالات زیر پاسخ دهید. (M ماتریس انتقال می‌باشد)

- Transition Probs: $P(S_1|S_1) = 0.8, P(S_2|S_1) = 0.2, P(S_1|S_2) = 0.2, P(S_2|S_2) = 0.8$
- Emission Probs: $P(A|S_1) = 0.4, P(C|S_1) = 0.1, P(G|S_1) = 0.4, P(T|S_1) = 0.1$
 $P(A|S_2) = 0.1, P(C|S_2) = 0.4, P(G|S_2) = 0.1, P(T|S_2) = 0.4$
- Start probabilities: $P(S_1) = P(S_2) = 0.5$

الف) (۲۰ نمره) $P(x|M)$ را با استفاده از الگوریتم forward محاسبه کنید.

ب) (۱۵ نمره) احتمال پسین $P(\pi_i = S_1|x, M)$ را برای مقادیر i از ۱ تا ۳ محاسبه کنید.

ج) (۱۵ نمره) محتمل‌ترین مسیر حالت‌های پنهان در الگوریتم Viterbi را به دست آورید.

Intro to ML (۵۰ نمره)

۱. (۵۰ نمره) دسته‌بندی متن، کاربردی است که در آن برای یک متن ارائه شده، دسته‌ای خاص از بین دسته‌های از پیش تعیین شده مشخص می‌شود. مدل بیز ساده‌انگارانه (Naive Bayes) ابزاری مناسب برای این کاربرد است و نوعاً از آن برای دسته‌بندی متن استفاده می‌شود. متغیر پرسش در این مسأله، دسته مربوط به متن است و متغیرهای effect، حضور یا عدم حضور هر کلمه از زبان مربوطه در متن است. فرض می‌شود که حضور یا عدم حضور کلمات در هر متن، از هم مستقل است و فراوانی حضور هر کلمه نیز بستگی به دسته متن (مثلاً ورزشی، سیاسی و ...) دارد.

- (۲۰ نمره) به طور دقیق تشریح کنید که مدل ساده‌انگارانه برای این مسأله چگونه ساخته می‌شود. فرض کنید مجموعه‌ای از متن‌های مختلف در اختیار داریم که دسته هر متن نیز مشخص شده است (مجموعه آموزشی).

- (۲۰ نمره) به طور دقیق تشریح کنید که برای یک متن جدید، چگونه می‌توان دسته آن را تشخیص داد.
- (۱۰ نمره) آیا فرض استقلال شرطی که مدل بیز ساده‌انگارانه بر آن استوار است، فرض منطقی‌ای در این مساله است؟ توضیح دهید.