



دانشکده مهندسی کامپیوتر

## تمرین چهارم مبانی یادگیری ماشین

پاییز 1402

## بخش تشریحی

### سوال اول

1. با الگوریتم آزادسازی نسبی محدودیت‌ها برای توسعه ماشین بردار پشتیبان برای حالت جدایی ناپذیر خطی آشنا شده‌اید. در این سوال با الگوریتمی مشابه آشنا می‌شوید که با مساله‌ی بهینه‌سازی زیر مطمح می‌شود:

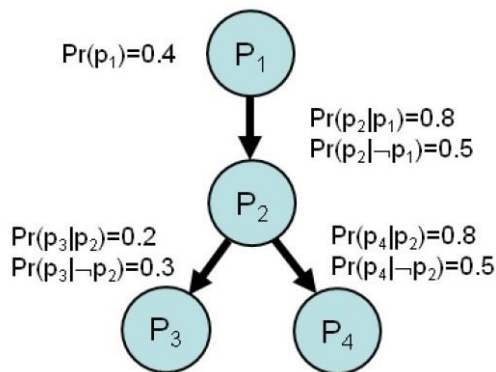
$$\min_{w,b,\xi} \quad \frac{1}{2} \|w\|^2 + \frac{C}{2} \sum_{i=1}^m \xi_i^2$$

$$\text{s.t.} \quad y^{(i)}(w^T x^{(i)} + b) \geq 1 - \xi_i, \quad i = 1, \dots, m$$

- (a) لاگرانژین این مساله‌ی بهینه‌سازی را پیدا کنید.
- (b) لاگرانژین به دست آمده از قسمت قبل را با استفاده از متغیرهای  $w, b$  و  $\xi$  کمینه کنید.
- (c) دوگان مساله را به دست آورید تا مساله به فرم بیشینه‌سازی تغییر یابد و آماده‌ی به دست آوردن ضرایب لاگرانژ گردد.

### سوال دوم

1. برای شبکه‌ی زیر  $\Pr(\neg p_3)$  و  $\Pr(p_2 | \neg p_3)$  را به کمک متد حذف متغیر به دست آورید.



## سوال سوم)

1. برای کاهش ابعاد یک مجموعه از داده‌ها به کمک متد تحلیل مولفه اصلی به ترتیب زیر عمل می‌کنیم:
- a. محاسبه‌ی ماتریس پراکندگی (کوواریانس) مجموعه داده.
  - b. محاسبه‌ی مقادیر ویژه و سپس بردارویژه‌ی ماتریس پراکندگی.
  - c. مرتب کردن مقادیر ویژه به دست آمده و بردارهای ویژه مربوط به آنها و انتخاب بردارویژه با بیشترین مقدار ویژه.
  - d. تبدیل داده‌ی اصلی به داده‌ی کاهش بعد یافته (مولفه‌ی اصلی) با ضرب ماتریس افکنش حاصل از بردارهای ویژه‌ی به دست آمده در ماتریس داده‌های اولیه.

با توجه به توضیحات داده شده، ابعاد مجموعه داده‌ی زیر را از 2 به 1 کاهش دهید و برای این کار از متد تحلیل مولفه اصلی استفاده نمایید.

Feature	Example 1	Example 2	Example 3	Example 4
X1	4	8	13	7
X2	11	4	5	14

## بخش پیاده سازی

### سوال اول)

در این سوال باید یک سیستم توصیه گر بر اساس فیلترسازی عصبی همکارانه (Neural Collaborative Filtering) برای مجموعه داده‌ی داده شده (Movie-lens) پیاده‌سازی کنید. کد مربوط به دانلود و آماده‌سازی مجموعه داده در نوت بوک محل تحویل قرار داده شده است. در این سوال باید مدل را پیاده‌سازی کرده و با استفاده از تابع زیر مناسب آن را آموزش دهید (درباره‌ی توابع مختلف مورد استفاده برای این مساله جست‌وجو کنید). اندازه‌ی جانشانی (Embedding) کاربران و فیلم‌ها را برابر با 64 در نظر بگیرید.

#### معرفی مجموعه داده:

مجموعه داده‌ی Movie-Lens شامل داده‌های مربوط به امتیازدهی کاربران به فیلم‌ها می‌باشد. در این سوال از نسخه‌ی صد هزار تایی این مجموعه داده (MovieLens 100K) استفاده می‌شود. همانگونه که از نام آن برمی‌آید، این مجموعه شامل یکصد هزار امتیازدهی است.

ستون‌های مجموعه داده:

- user\_id: شناسه‌ی یکتای کاربر.
- item\_id: شناسه‌ی یکتای فیلم.
- rating: امتیاز داده شده به فیلم مذکور توسط کاربر که رنج 1 (ضعیف) تا 5 (عالی) را شامل می‌شود.
- timeStamp: تاریخ و زمانی که امتیاز توسط کاربر ثبت شده است.

در خروجی، نمودار تغییرات Loss، معیار Recall@10 و نام و مشخصات ده فیلم پیشنهاد شده برای یک کاربر دلخواه را نشان دهید.

## سوال دوم)

در این سوال باید با استفاده از روش XGBoost، مشخص کنید که هر بیمار، بیماری قلبی دارد یا خیر. 80 درصد داده ها را برای آموزش در نظر بگیرید و معیار Accuracy را برای داده های آموزش و آزمایش گزارش کنید.

### معرفی مجموعه داده:

مجموعه داده شامل ویژگی های یک بیمار و یک متغیر هدف است که دو مقدار 0 و 1 به معنی وجود و عدم وجود بیماری قلبی دارد.

از لینک زیر میتوانید به مجموعه داده دسترسی داشته باشید و جزییات آن را مشاهده کنید.

[Heart Disease - UCI Machine Learning Repository](https://archive.ics.uci.edu/ml/datasets/Heart+Disease)

برای هر کدام از هایپرپارامتر های `n_estimators`, `max_depth`, `learning_rate`، دو مقدار متفاوت را آزمایش کرده و نتایج را گزارش دهید.

---

همه سوالات را بر روی تمرین گیت هاب با آدرس

<https://classroom.github.com/a/9unBqbRi>

تحويل دهید.

آخرین مهلت تحويل تمرينات، جمعه 29 دی میباشد.

موفق باشید.