



**دانشگاه صنعتی امیر کبیر**  
(پلی تکنیک تهران)

به نام خدا

تمرین ششم درس هوش مصنوعی

عنوان: Bayesian Network

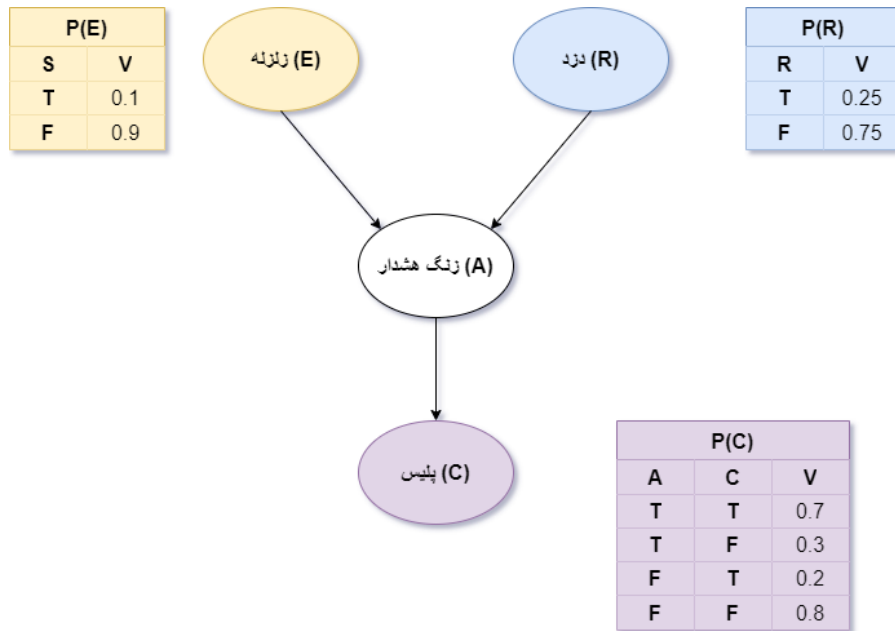
استاد درس:

دکتر اکبری

موعده تحویل: 1401/03/17

## سوالات تشریحی

1. شکل زیر به شبکه Bayesian را نمایش میدهد که مربوط به دستگاهی است که اگر دزد وارد خانه شود یا زلزله بیاید به صدا در میاد و مستقیماً به پاسگاه پلیس وصل است و آن ها را خبر میکند و پلیس به سمت خانه نیرو میفرستد . این دستگاه از آنجایی که تولید وطنی است، دقت بسیار بالایی دارد و خطای آن صفر است.



حال از شما میخواهیم که به سوالات زیر جواب دهید:

$$P(C=T, A=T, E=T, R=T) \quad (i)$$

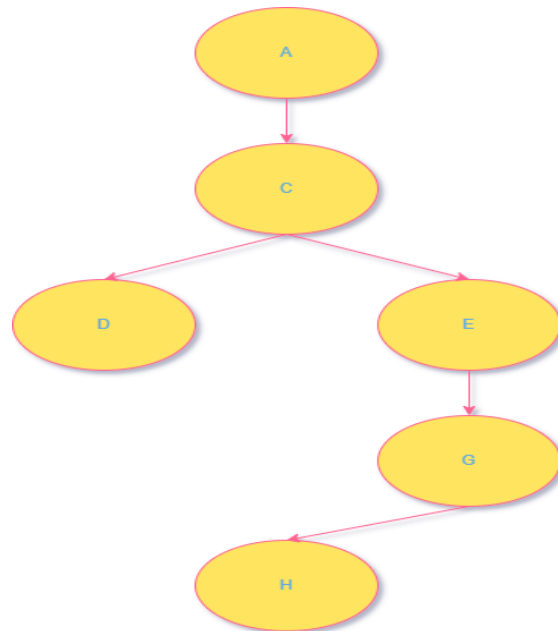
$$P(C=F, A=T, E=F, R=T) \quad (ii)$$

$$P(C=T, A=T, R=T) \quad (iii)$$

توجه داشته باشید که حتماً راه حل خود را نیز بنویسید و تنها به نوشتن به عدد اکتفا نکنید.

2. شخص بیماری به دکتر مراجعه میکند، دکتر از آنجایی که سواد بالایی در تشخیص بیماری دارد باید بفهمد که شخص کرونا دارد یا بیماری او نشان از آنفولانزا است . او می داند که نشانه های آنفولانزا سرفه و تب و نشانه های کرونا تنگی نفس و تب است. حال از شما دانشجویان میخواهیم که ابتدا متغیر های مسئله را نام ببرید و سپس شبکه Bayesian تشخیص بیماری را رسم کنید و در آخر فرمول کلی احتمال را بنویسید.

3. با استفاده از شبکه زیر به سوالات پاسخ دهید و دلیل پاسخ خود را بنویسید.



- I. آیا E و D لزوماً از هم مستقل اند با توجه به اینکه A داده بشود؟
- II. آیا A و C لزوماً از هم مستقل اند با توجه به اینکه D داده بشود؟
- III. آیا A و H لزوماً از هم مستقل اند با توجه به اینکه C داده بشود؟

4. برای تشخیص ایمیل‌های اسپم می‌خواهیم از یک مدل naïve bayes استفاده کنیم. برای ساخت جدول احتمالات از 4 نمونه داده آموزشی استفاده می‌شود. در آخر به کمک جدول‌های بدست آمده درباره لیبل داده تست تصمیم‌گیری می‌شود. (در هنگام محاسبه احتمالات در نظر داشته باشید که به تعداد تکرار هر کلمه 1 واحد اضافه کنید تا در هنگام محاسبه احتمال شرطی وجود احتمال صفر، باعث صفر شدن حاصل ضرب نشود)

الف) جدول احتمالات را برای لیبل‌های ham, spam محاسبه کنید.

ب) با کمک جدول‌ها لیبل داده تست را تعیین کنید.

	Doc	Words	Class
Training	1	میلیارد فوری فرصت برنده	spam
	2	فوری فوری تخفیف برنده	spam
	3	جلسه فردا تنظیم	ham
	4	فوری جلسه فردا خبر	ham
Test	5	تخفیف آموزش میلیارد برنده	؟

## پیاده‌سازی

ساخت مدل naïve bayes برای تشخیص ایمیل‌های اسپم.

**توضیحات:** ساخت این مدل دقیقاً شبیه تمرین آخر بخش تشریحی است. با این تفاوت که باید الگوریتم naïve bayes را به صورت کد بنویسید.

داده‌ای که برای آموزش در نظر گرفته شده شامل 300 ایمیل spam و 300 ایمیل ham(nonsпам) است. و 400 نمونه ایمیل نیز برای بخش تست مدل ساخته شده مورد استفاده قرار میگیرد.

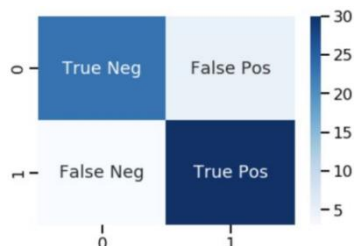
**نکته:** تمام مراحل پیش‌پردازش داده شامل خواندن متن ایمیل‌ها از فایل، حذف کلمات توقف، Tokenizing، Vectorizing و انتخاب فیچرهای برتر جهت آموزش مدل برای شما آماده شده است. صرفاً در ابتدای کار باید کدهای مربوطه توسط شما اجرا شوند.

در انتهای مراحل پیش‌پردازش، به شما لیستی داده میشود که شامل وکتورهای 600 ایمیل مرحله training است. هر یک از این وکتورها دارای 700 بعد میباشد که برابر با تعداد فیچرهای برتر انتخاب شده در مرحله قبلیست. مقدار هر اندیس نشان دهنده تعداد تکرار آن فیچر (کلمه) در آن سمپل است.

- شما با کمک وکتورها میتوانید جدول‌های احتمالات را بسازید (این کار دقیقاً مانند مرحله الف سوال چهارم بخش تشریحی است. با این تفاوت که تعداد تکرار هر کلمه در هر سمپل نیز آماده شده است).
- در notebook هم توضیحاتی قرار داده شده، ویدیو آموزشی 15 دقیقه‌ای که در پیاده‌سازی الگوریتم به شما کمک میکند.
- لینک: [colab notebook](#) که بعد از باز کردن باید یک کپی از آن بسازید که داخل google drive تان ذخیره شود. بعد از اضافه کردن کد خود، در colab notebook روی share کلیک کرده و notebook را در حالت Viewer for any one قرار دهید. سپس لینک notebook خود را در گزارش کار یا در یک فایل txt ذکر کنید.
- توصیه میشود از colab notebook استفاده شود در غیر این صورت میتوانید از ipynb که در فایل تمرین قرار دارد استفاده کرده و آن را در محیط jupyter باز کنید. (فیلم آموزشی 3 دقیقه‌ای نیز برای کار با colab notebook برای شما قرار داده شده است).

### بخش امتیازی:

- در پایان ساخت مدل قطعه کدی جهت سنجش دقت مدل شما قرار داده شده است؛ معیارهای مورد استفاده در آن، F1, precision, recall اند. در مورد مزیت و معایب هر یک از این معیارها تحقیق کنید و نتایج خود را بنویسید.
- ماتریس در هم ریختگی (confusion matrix) را برای ارزیابی مدل خود رسم کنید. (برای بدست آوردن confusion matrix میتوانید از کتابخانه scikit-learn استفاده کنید).



به عنوان نمونه:

### نکات مهم

- جهت انجام بخش پیاده‌سازی، از Jupyter یا colab استفاده کنید و فایل نهایی را با پسوند ipynb آپلود کنید. همچنین شماره دانشجویی خود را به عنوان نام فایل در نظر بگیرید.
- استفاده از قطعه کدهای آماده برای پیاده‌سازی الگوریتم مجاز نیست.

- برای بخش پیاده سازی صرفا کافی است در notebook توضیحات کد های خود را بنویسید، برای راحتی کار میتوانید در آخر پیاده سازی سلولی باز کرده و توضیحات را در آنجا قرار دهید.
- هرگونه شباهت در گزارش و پاسخ تشریحی به منزله تقلب می باشد و کل نمره تمرین صفر می باشد. ( می توانید از اینترنت به عنوان منبع کمکی هم در سوالات تشریحی و هم در سوالات پیاده سازی استفاده کنید، اما منبع آن را ذکر کنید و کپی برداری ممنوع می باشد و نمره ی صفر تعلق می گیرد)
- فایل pdf و کد را بصورت یکجا در قالب یک فایل zip در سامانه [courses](#) آپلود کنید (نام فایل = شماره دانشجویی).

## ارتباط با ما

جهت مطرح کردن سوالات و ابهام هایی که دارید می توانید از طریق تلگرام یا ایمیل با ما در ارتباط باشید.

@amir2647j

امیرحسین جغتائی

erfansamieyan.s@gmail.com

عرفان سمیعیان