

بوت کمپ یادگیری ماشین



بوت کمپ
BOOTCAMP
یادگیری ماشین
MACHINE
LEARNING
مهلت ثبت نام تا ۱۷ دی

قبل از همه چیز، آیا دوره مناسب شماست؟

احتمالا متوجه شدید که این دوره قراره تخصصی باشه و برای این که بتونید از دوره استفاده بکنید، لازمه حتما دانش ماشین لرنینگ خوبی رو از قبل داشته باشید. این دوره مناسب کسانیه که دانش تئوری رو خودشون کسب کردن و الان دنبال پروژه زدن و کار عملی هستن. برای اینکه بهتر بتونین متوجه بشید که آیا دوره مناسب شما هست یا نه، پیشنهاد می‌کنیم سوال‌های زیر رو با بله یا خیر جواب بدید.

۱- می‌تونم بدون استفاده از اینترنت یک linear regression را پیاده‌سازی کنم.

۲- می‌تونم یک فرضیه را با استفاده از روش‌های آماری قبول یا رد کنم.

۳- با الگوریتم gradient descent و جزئیات آن آشنایی دارم.

۴- با MLE و جزئیات ریاضی آن آشنایی دارم.

۵- حداقل سه مسئله را با استفاده از روش‌های حوزه ماشین‌لرنینگ از صفر تا صد پیاده‌سازی کرده‌ام.

۶- با اعمال مربوط به ماتریس‌ها مثل جمع، ضرب، ترانزپوز و امثالهم آشنایی دارم و می‌توانم eigenvalue و eigenvector ها را برای یک ماتریس بدست آورم.

۷- با مفاهیم loss function و optimizer شبکه عصبی آشنایی دارم.

۸- با زبان برنامه‌نویسی پایتون تا حد خوبی آشنایی دارم و کار کردم.

۹- از ابزارهای visualization پایتون استفاده کردم.

۱۰- با عموم الگوریتم‌های ماشین‌لرنینگ مثل K means، KNN، linear regression، svm، decision tree و ... آشنا هستم و طرز کار هر کدام را بلدم.

۱۱- با spark آشنایی دارم.

۱۲- با bayes rule آشنایی دارم.

۱۳- حداقل یک بار از کتابخانه keras استفاده کرده‌ام.

۱۴- فرق data scientist و data analyst را می‌دانم.

حالا تعداد جواب‌های بله خود را بشمارید.

● اگر به بیش از ۱۰ سوال جواب مثبت دادید، دوره بوت‌کمپ مناسب‌تونه و می‌تونید در اولین فرصت چالش رو انجام بدید و فرایند ثبت‌نامتون رو تکمیل کنید. البته که گذروندن آموزش‌های قرار گرفته در بخش Warm Up و مطالعه درباره مواردی که پاسخ منفی بهشون دادید کمک می‌کنه که خیالتون بابت اینکه از دوره بیشترین استفاده رو ببرید راحت بشه.

● اگر تعداد جواب‌های مثبت شما کمتر از ۱۰ تاست نیازه که دقیق‌تر وضعیت خودتون رو ارزیابی کنید و حتما تا قبل از شروع بوت‌کمپ یه سر به دوره‌های مقدماتی و پیشرفته ماشین لرنینگ بزنید. اگه اول راهید پیشنهاد می‌کنیم دوره مقدماتی ماشین لرنینگ رو بگذرونید تا با مباحث اولیه و پایه‌ای این حوزه آشنا بشید و حین انجام پروژه یکم دستتون به داده کثیف بشه. اگر که تعداد جواب‌های مثبت‌تون نزدیک ۱۰ بود یا با اغماض به ۱۰ تا می‌رسید می‌تونین دوره پیشرفته ماشین لرنینگ رو بگذرونید تا کامل برای بوت‌کمپ آماده بشید. توی این دوره مباحث بیشتری با عمق بیشتر توضیح داده شده که می‌تونه کمکتون کنه.

قبل از همه چیز، آیا دوره مناسب شماست؟

۲. چالش ورود به بوت کمپ

۳. گرم بشیم برای یادگیری!

۱. قبل از همه چیز، آیا دوره مناسب شماست؟

۲. چالش ورود به بوت کمپ

۳. گرم بشیم برای یادگیری!

و حالا، این شما و این چالش!

اگر بعد از جواب دادن به سوالات خودارزیابی، تصمیم گرفتید با به سری مطالعات که توی زمان باقی مونده تا دوره دارید خودتون رو برای شرکت در بوت کمپ آماده کنید، پس الان موقع اینه که چالش زیر رو حل کنید و جوابتون رو برای ما ارسال کنید.

Twitch به کاربرانش این امکان رو میده که یک کانال بسازن و بتونن به صورت تعاملی با اعضای کانالشون یک livestream داشته باشن. شما اخیرا به عنوان data scientist توی شرکت twitch استخدام شدین. یکی از مشکلاتی که توی شرکت مطرح شده اینه که کاربرهای twitch دارن علاقشون رو به این پلتفرم از دست میدن و نمیتونن کانالهایی رو پیدا کنن که مورد علاقشون باشه. برای اینکه به کاربرها کمک کنین شما پیشنهاد می‌دین که یک recommender system پیاده‌سازی کنیم که بشه با استفاده از اون به کاربر یک کانال جدید پیشنهاد بدیم که مورد علاقه‌اش باشه (دقت کنین که این کانال حتما باید جدید باشه وگرنه خیلی هنری نکردیم).

دیتاستی که در اختیارتون قرار داده میشه توی این [لینک](#) قابل مشاهده است. این دیتاست توی یک بازه ۴۳ روزه جمع‌آوری شده. هر ۱۰ دقیقه از همه stream ها یک snapshot گرفته می‌شه و داده‌اش ذخیره می‌شه. البته ما قراره روی به قسمتی از این داده کار کنیم. دیتا فقط شامل کاربرانی می‌باشد که داخل چت فعالیت داشتند. یعنی لزوما تمامی بینندگان stream ها مشاهده نمی‌شه. دیتا ۵ تا ویژگی داره:

- User_id : آیدی مربوط به کاربر
- Stream_id: آیدی مربوط به stream ای که کاربر در قسمت چت فعالیت داشته
- Streamer_name: نام کانال مربوط به stream
- start_time: شماره اولین snapshot ای که کاربر چت stream مربوط به stream_id فعالیت داشته
- Finish_time: شماره آخرین snapshot ای که کاربر چت stream مربوط به stream_id فعالیت داشته

توی این [لینک](#) می‌تونین توضیحات بیشتری درباره دیتاست پیدا کنین.

دقت کنین که اولین کاری که به data scientist وقتی به به مسئله برمی‌خوره انجام میده مدل زدن نیست. نگاه کردن و آنالیز کردن دیتا می‌تونه خیلی به فهم مسئله کمک کنه. بعد از اینکه مسئله رو فهمیدین به متریک ارزیابی درست انتخاب کنین و بر اساس اون روشتون رو ارزیابی کنین.

دقت کنید:

- حتما داخل محیط [google colab](#) پیاده‌سازی رو انجام بدید و کد رو کامل اجرا کنید و در نهایت لینک کدتون رو پابلیک کنید و لینکش رو با ما به اشتراک بذارید.
- توی این چالش ارزیابی شما بر اساس دقتی که می‌گیرید نیست. مهمه که چجوری با مسئله برخورد می‌کنین و چجوری روش فکر می‌کنین و چجوری کد می‌زنین. برای همین حتما سعی کنید با documentation خوب به ما کمک کنین که راحت‌تر متوجه روشتون بشیم.
- مستندات مربوطه رو داخل خود کد بنویسید. یا به صورت کامنت یا به صورت markdown.

۱. قبل از همه چیز،
آیا دوره مناسب شماست؟

۲. چالش ورود به بوت کمپ

۳. گرم بشیم برای یادگیری!

گرم بشیم برای یادگیری!

یکی از مهم‌ترین کارهایی که هر فرد لازمه در حوزه تخصصیش انجام بده، مطالعه پیوسته مطالب اون حوزه و به روز بودن. یه تعداد محتوای کاربردی رو برای مطالعه این زیر بهتون معرفی کردیم. پیشنهاد می‌کنیم توی مدتی که تا شروع بوت‌کمپ فرصت دارید این منابع رو مطالعه کنید تا هم برای پاسخ به چالش ورودی بتونید ازشون کمک بگیرید و هم در صورت پذیرش نهایی در بوت‌کمپ جعبه ابزار مهارت‌هاتون برای انجام پروژه تکمیل‌تر باشه.

الگوریتم‌های پرکاربرد ماشین‌لرنینگ رو لازمه بدونین پشت‌شون چه اتفاقی میفته. می‌تونین برای این کار اسمشون رو سرچ کنین و یاد بگیرین هر کدوم چجوری کار می‌کنن ولی اگه می‌خواین عمیق‌تر یاد بگیرین کتاب hands on machine learning قسمت ماشین‌لرنینگش رو بخونین.چندتا نمونه از الگوریتم‌های پرکاربرد اینان:

- Linear regression
- Logistic regression
- Decision tree
- Random forest
- Svm
- Naive Bayes
- KNN
- Kmeans
- Ensemble learning

۲

برای یادگیری deep learning باز هم کتاب hands on machine learning قسمت دیپ‌لرنینگش خیلی توصیه می‌شه ولی اگه تا حالا اسم deep learning رو هم نشنیدین بنظرم **دوره رایگان Andrew ng** می‌تونه کمکتون کنه.

۳

برای این‌که آمار لازم برای کار ماشین‌لرنینگ رو یاد بگیرین کتاب practical statistics for data scientists کتاب خوبیه. پیشنهاد می‌کنیم اگه وقت خوندن کل کتاب رو ندارین فصل‌های زیر رو مطالعه کنید:

- EDA
- Sampling
- Distributions

۴

برای آشنا شدن با جبر خطی ویدئوهای این **کانال یوتیوب** رو می‌تونین ببینین. خیلی مقدماتی توضیح می‌ده ولی دستتون میاد چیا لازمه یاد بگیرین.

۵

کد زدن با زبان پایتون و آشنایی با کتابخانه‌های حوزه ماشین‌لرنینگ برای این دوره لازمه. منابع خوبی برای یادگیری این مطالب توی اینترنت هست و هر منبعی که باهاش راحت بودین رو انتخاب کنین. ولی خب مثلا برای pandas این **دوره کگل** می‌تونه کمکتون کنه و برای ویژوالیزیشن این دوره.

۶

مطالبی که در بالا گفته شد توی **دوره‌های پیشرفته و مقدماتی ماشین‌لرنینگ** رهنما کالج به صورت مفصل پرداخته شده. گذروندن اون دوره‌ها می‌تونه شروع خوبی برای آماده شدن تون برای بوت‌کمپ باشه.

امیدواریم بین نفرات نهایی پذیرفته شده در دوره بینیمتون!



RAHNEMA COLLEGE

Hang out with the professionals!