

# جزوه دوره

جامپ یادگیری ماشین College Quera

تهیه شده توسط علی فاضل نیا

دانشجوی علوم کامپیوتر

---

راههای ارتباط:

Email: alifazelniya.1384@gmail.com

Telegram: @Norbert\_desu

GitHub: github.com/AliFazelniya

# فهرست مطالب

ت	پیشگفتار
۱	۱ مقدمه
۱	۱.۱ سلام!
۲	۲.۱ قالب کالج
۳	۳.۱ سیستم امتیازات و دریافت گواهی اصلی
۳	۴.۱ یادگیری ماشین چیست؟
۳	۵.۱ چرا پایتون؟
۳	۶.۱ آماده‌سازی محیط کار
۴	۷.۱ اجرای نوت‌بوک‌ها
۴	۸.۱ گوگل کولب
۴	۹.۱ معرفی مجموعه‌داده
۵	۲ مدیریت پروژه
۵	۱.۲ اهداف فصل
۵	۲.۲ چرخه پروژه
۵	۳.۲ اولویت‌بندی
۵	۴.۲ سازماندهی تیم
۵	۵.۲ چرا پروژه‌ها شکست می‌خورند؟
۶	۳ آماده‌سازی داده
۶	۱.۳ اهداف فصل
۶	۲.۳ سوالاتی درباره داده
۶	۳.۳ چالش‌های داده
۶	۴.۳ ویژگی‌های مجموعه‌داده‌ی خوب
۶	۵.۳ تقسیم‌بندی مجموعه‌داده
۷	۶.۳ داده‌های پرت
۷	۷.۳ مقادیر گم‌شده
۷	۸.۳ مجموعه‌داده نامتوازن
۸	۴ مهندسی ویژگی
۸	۱.۴ اهمیت
۸	۲.۴ مشخصات ویژگی خوب

۸	ویژگی‌های دسته‌ای	۳.۴
۸	مقادیر گم شده در ویژگی‌های دسته‌ای	۴.۴
۸	ویژگی‌های تقویمی	۵.۴
۸	سنتر ویژگی	۶.۴
۹	تغییر مقیاس ویژگی	۷.۴
۹	نشت داده	۸.۴
۹	فوت و فن‌های مهندسی ویژگی	۹.۴
۹	کاهش ابعاد	۱۰.۴
۹	انتخاب ویژگی	۱۱.۴
۹	خط لوله	۱۲.۴
۱۰	<b>رگرسیون</b>	<b>۵</b>
۱۰	اهداف فصل	۱.۵
۱۰	مقدمه	۲.۵
۱۰	مدل چیست؟	۳.۵
۱۰	تخمین، تابع هزینه و بهینه‌سازی	۴.۵
۱۰	رگرسیون خطی	۵.۵
۱۰	ارزیابی	۶.۵
۱۱	رگرسیون چندجمله‌ای	۷.۵
۱۱	عمومیت	۸.۵
۱۱	رگولاریزیشن	۹.۵
۱۲	<b>دسته‌بندی</b>	<b>۶</b>
۱۲	مقدمه	۱.۶
۱۲	رگرسیون لجستیک	۲.۶
۱۲	ارزیابی - قسمت اول	۳.۶
۱۲	ارزیابی - قسمت دوم	۴.۶
۱۲	کراس ولیدیشن	۵.۶
۱۲	نزدیک‌ترین- $k$ همسایه	۶.۶
۱۳	بیز ساده‌لوحانه	۷.۶
۱۳	ماشین بردار پشتیبان	۸.۶
۱۳	هایپرپارامترها	۹.۶
۱۳	آشنایی با کتابخانه $\text{O}2\text{H}$	۱۰.۶
۱۳	درخت تصمیم	۱۱.۶
۱۳	فوت و فن درخت تصمیم	۱۲.۶
۱۳	بیش‌برازش درخت تصمیم	۱۳.۶
۱۴	<b>یادگیری تجمعی</b>	<b>۷</b>
۱۴	اهداف فصل	۱.۷
۱۴	مقدمه	۲.۷
۱۴	جنگل تصادفی	۳.۷

۱۴	.....	الگوریتم AdaBoost	۴.۷
۱۴	.....	الگوریتم Boosting Gradient	۵.۷
۱۴	.....	الگوریتم XGboost	۶.۷
۱۵	.....	روش Stacking	۷.۷
۱۶	.....	<b>پروژه اول</b>	<b>۸</b>
۱۶	.....	۱.۸ مقدمه	
۱۶	.....	۲.۸ یادداشت‌ها و راه حل	
۱۷	.....	<b>شبکه عصبی</b>	<b>۹</b>
۱۷	.....	۱.۹ اهداف فصل	
۱۷	.....	۲.۹ پرسپیترون	
۱۷	.....	۳.۹ آموزش پرسپیترون	
۱۷	.....	۴.۹ پرسپیترون چندلایه	
۱۷	.....	۵.۹ عمومیت	
۱۸	.....	<b>یادگیری ناظارت‌نشده</b>	<b>۱۰</b>
۱۸	.....	۱.۱۰ مقدمه	
۱۸	.....	۲.۱۰ الگوریتم PCA	
۱۸	.....	۳.۱۰ الگوریتم t-SNE	
۱۸	.....	۴.۱۰ خوشبندی با k-means	
۱۸	.....	۵.۱۰ خوشبندی با k-modes	
۱۸	.....	۶.۱۰ خوشبندی با k-prototype	
۱۹	.....	<b>پروژه دوم</b>	<b>۱۱</b>
۱۹	.....	۱.۱۱ اهداف فصل	
۱۹	.....	۲.۱۱ تعبیه‌ی متن	
۱۹	.....	۳.۱۱ فاصله‌ی ویرایش	
۱۹	.....	۴.۱۱ معیار شباهت RBO	
۲۰	.....	<b>بیشتر بدانید</b>	<b>۱۲</b>
۲۰	.....	۱.۱۲ نمونه‌کاهی با NearMiss	
۲۰	.....	۲.۱۲ نمونه‌افزایی با SMOTE	
۲۰	.....	۳.۱۲ درخت رگرسیون	
۲۱	.....	<b>واژه‌نامه</b>	
۲۲	.....	<b>منابع</b>	

# پیشگفتار

اینجا هدف، دامنه، و نحوه استفاده از جزو را بنویس. (می‌توانی همین متن را بعداً دقیقاً با متن خودت جایگزین کنی.)

# فصل ۱

## مقدمه

### ۱.۱ سلام!

سلام؛ ورود شما را به کالج «یادگیری ماشین ۲ | جامپ تکنیکال» خوشآمد می‌گوییم. از اینکه در کوئرا کالج افتخار میزبانی شما را داریم، به خود می‌بالیم

«یادگیری ماشین ۲ | جامپ تکنیکال» کالج سوم از مسیر علم داده و یادگیری ماشین کوئراست که پس از «یادگیری ماشین ۰ | دروازه ورود» و «یادگیری ماشین ۱ | تحلیل داده با پایتون» طراحی شده است.

هدف ما از تدوین کالج این است که شما را به شکل اصولی و گامبهگام با الگوریتم‌های یادگیری ماشین آشنا کنیم؛ به‌طوری که در نهایت بتوانید الگوریتم‌ها را تحلیل کنید، نقاط ضعف و قوت آن‌ها را بشناسید و تشخیص دهید چگونه از آن‌ها برای حل مسائل دنیای واقعی کمک بگیرید! علاوه بر این‌که با الگوریتم‌های یادگیری ماشین آشنا می‌شویم، تکنیک‌هایی برای استفاده عملی از آن‌ها را نیز خواهید آموخت؛ به عبارت بهتر، «یادگیری ماشین ۲ | جامپ تکنیکال» آمیخته‌ای موزون از دانش علمی و مهارت عملی یادگیری ماشین کلاسیک است!

این کالج مناسب افرادی است که در حد متوسط با پایتون و کتابخانه‌های Numpy و Pandas آشنایی داشته باشند. داشتن دانش حداقلی از آمار احتمال و حسابان (مباحث مشتق) و جبر خطی (ماتریس‌ها) به شما در فراگیری محتوای این کالج کمک می‌کند. به صورت کلی می‌توان گفت اگر با ریاضیات در حد مقطع دبیرستان آشنا باشید، به راحتی می‌توانید به ماجراجویی در این کالج بپردازید! اگر با پیش‌نیازهای کالج آشنا نیستید، می‌توانید با گذراندن سایر دوره‌های کوئرا کالج، مهارت‌های لازم را کسب کنید.

شما با داشتن پیش‌نیازهای کالج شروع می‌کنید و در انتهای پس از مطالعه درسنامه‌ها و حل تمرین‌ها، به «دانشمند داده» و «مهندس یادگیری ماشین» تبدیل می‌شوید که قطعاً راه طولانی اما شیرینی برای حرفه‌ای شدن پیش رو دارید!

البته برای حفظ تناسب کالج و افزایش احتمال یادگیری عمیق، از آموزش کامل یادگیری عمیق، که زیرمجموعه یادگیری ماشین است، پرهیز شده است؛ بلکه شبکه‌های عصبی مصنوعی و مقدمات یادگیری عمیق را در فصل «شبکه عصبی» آورده‌ایم. برای فراگیری یادگیری عمیق، می‌توانید منتظر کالج‌های بعدی مسیر علم داده و یادگیری ماشین کوئرا کالج باشید

امیدواریم با ارائه‌ی آموزش باکیفیت، گامی مثبت در افزایش دانش و پیشرفت شما برداریم. سختی‌های

مسیر نه تنها ناامیدمان نمی‌کند بلکه توانمان را بیشتر و تصمیم‌مان را راسخ‌تر می‌کند. مسیر رسیدن به قله‌ای که آرزویش را داریم؛ جوانانی توانمند، پرتلash و ایرانی پیشرفته...



## ۲.۱ قالب کالج

### قالب کالج روند آموزشی کوئرا کالج

امروزه با فرآگیر شدن آموزش‌های آنلاین، قالب‌های متعددی برای یادگیری مفاهیم علوم کامپیوتر بصورت آنلاین پیاده‌سازی شده است. آموزش آنلاین این مفاهیم، فرصت برنامه‌نویسی به همراه یادگیری را فراهم می‌کند و همچنین انجام تمرین‌های واقعی و استفاده از کتابخانه‌ها و قالب‌های نزدیک به صنعت را تسهیل می‌کند. اما آموزش آنلاین چالش‌هایی نیز به همراه دارد؛ زیرا تعاملی که استاد در سر کلاس با دانشجو دارد و نظمی که جلسات کلاس به یادگیری دانشجو می‌دهد به سختی در قالب‌های آموزش آنلاین گنجانده می‌شود. در کوئرا کالج تلاش کردیم قالبی برای برطرف شدن نیاز تعاملی بودن، نظم، و همچنین عملی بودن آموزش آماده کنیم.

در کل، آموزش این دوره متشکل از چندین فصل می‌باشد، که هر فصل شامل تعدادی درسنامه، تمرین یا آزمون است.

#### درسنامه‌ها

پس از مطالعه‌ی هر درسنامه می‌توانید تیک "خواندم" آن را بزنید و پیش بروید. در پایین هر درسنامه بخش کامنت‌ها تعبیه شده تا بتوانید پرسش‌ها و نظرات عمومی خود را درباره‌ی درسنامه‌ی مربوطه با ما و سایر

شرکت‌کنندگان دوره مطرح کنید.

#### تمرین‌ها

تمرین‌ها نیز همگی توسط سامانه‌ی داوری خودکار Quera تصحیح شده، و پس از ارسال کد، توسط سیستم، نمره‌دهی می‌شود. داوری تمرین ممکن است شامل چندین تست مختلف باشد و سعی شده هنگامیکه در کسب نمره‌ی کامل یک تست دچار مشکل می‌شود بازخورد مناسبی توسط سیستم خروجی داده شود.

در تمرین‌هایی که عملکرد مدل یادگیری ماشین شما سنجیده می‌شود یک حد آستانه (Threshold) تعريف شده و در صورتی‌که مدل شما عملکردی بهتر از آن حد داشته باشد تمرین با موفقیت گذرانده می‌شود. البته اگر عملکرد مدل شما از مقدار خواسته‌شده بهتر باشد امتیاز اضافه‌تری کسب خواهد کرد.

#### آزمون‌ها

در بعضی از فصل‌ها به منظور درک عمیق‌تر مباحث آموزه شده تعدادی آزمون چندگزینه‌ای در نظر گرفته شده است. پاسخ شما به آزمون‌ها نیز مشابه با تمرین‌ها توسط سامانه‌ی داوری خودکار Quera انجام خواهد گرفت. هرچند که از نظر تعداد ارسال پاسخ محدودیتی نخواهید داشت، با این حال امتیاز آزمون‌ها در مقایسه با تمرین‌ها کمتر در نظر گرفته شده است.

#### پشتیبانی

ما گام به گام در طول مسیر این کالج همراهتان هستیم! در صورت وجود هرگونه پرسش یا ابهام درباره‌ی هر بخشی از این کالج می‌توانید از طریق پیغام خصوصی با پشتیبانان کالج در ارتباط باشید.

## ۳. سیستم امتیازات و دریافت گواهی اصلی

[ ]

## ۴. یادگیری ماشین چیست؟

[ ]

## ۵. چرا پایتون؟

[ ]

## ۶. آماده‌سازی محیط کار

[ ]

## ۷.۱ اجرای نوتبوک‌ها

] [

### ۸.۱ گوگل کولب

] [

### ۹.۱ معرفی مجموعه داده

] [

## فصل ۲

### مدیریت پروژه

#### ۱.۲ اهداف فصل

] [

#### ۲.۲ چرخه پروژه

] [

#### ۳.۲ اولویت‌بندی

] [

#### ۴.۲ سازماندهی تیم

] [

#### ۵.۲ چرا پروژه‌ها شکست می‌خورند؟

] [

## **فصل ۳**

### **آماده‌سازی داده**

#### **۱.۳ اهداف فصل**

] [

#### **۲.۳ سوالاتی درباره داده**

] [

#### **۳.۳ چالش‌های داده**

] [

#### **۴.۳ ویژگی‌های مجموعه‌داده‌ی خوب**

] [

#### **۵.۳ تقسیم‌بندی مجموعه‌داده**

] [

## ۶.۳ داده‌های پرت

] [

## ۷.۳ مقادیر گم شده

] [

## ۸.۳ مجموعه داده نامتوازن

] [

## فصل ۴

### مهندسی ویژگی

#### ۱.۴ اهمیت

] [

#### ۲.۴ مشخصات ویژگی خوب

] [

#### ۳.۴ ویژگی‌های دسته‌ای

] [

#### ۴.۴ مقادیر گم شده در ویژگی‌های دسته‌ای

] [

#### ۵.۴ ویژگی‌های تقویمی

] [

#### ۶.۴ سنتز ویژگی

] [

## ۷.۴ تغییر مقیاس ویژگی

] [

## ۸.۴ نشت داده

] [

## ۹.۴ فوت و فن‌های مهندسی ویژگی

] [

## ۱۰.۴ کاهش ابعاد

] [

## ۱۱.۴ انتخاب ویژگی

] [

## ۱۲.۴ خط لوله

] [

## فصل ۵

### رگرسیون

#### ۱.۵ اهداف فصل

] [

#### ۲.۵ مقدمه

] [

#### ۳.۵ مدل چیست؟

] [

#### ۴.۵ تخمین، تابع هزینه و بهینه‌سازی

] [

#### ۵.۵ رگرسیون خطی

] [

#### ۶.۵ ارزیابی

] [

## ۷.۵ رگرسیون چندجمله‌ای

] [

## ۸.۵ عمومیت

] [

## ۹.۵ رگولاریزیشن

] [

## فصل ۶

### دسته‌بندی

#### ۱.۶ مقدمه

] [

#### ۲.۶ رگرسیون لجستیک

] [

#### ۳.۶ ارزیابی - قسمت اول

] [

#### ۴.۶ ارزیابی - قسمت دوم

] [

#### ۵.۶ کراس ولیدیشن

] [

#### ۶.۶ نزدیک‌ترین- $k$ -همسایه

] [

## ۷.۶ بیز ساده لوحانه

] [

## ۸.۶ ماشین بردار پشتیبان

] [

## ۹.۶ هایپرپارامترها

] [

## ۱۰.۶ آشنایی با کتابخانهی $\text{O}2\text{H}$

] [

## ۱۱.۶ درخت تصمیم

] [

## ۱۲.۶ فوت و فن درخت تصمیم

] [

## ۱۳.۶ بیشبرازش درخت تصمیم

] [

## فصل ٧

### يادگیری تجمعی

#### ١.٧ اهداف فصل

] [

#### ٢.٧ مقدمه

] [

#### ٣.٧ جنگل تصادفی

] [

#### ٤.٧ الگوريتم AdaBoost

] [

#### ٥.٧ الگوريتم Boosting Gradient

] [

#### ٦.٧ الگوريتم XGboost

] [

## Stacking روش ۷.۷

]

## فصل ۸

### پروژه اول

۱.۸ مقدمه

] [

۲.۸ یادداشت‌ها و راه حل

] [

# فصل ۹

## شبکه عصبی

### ۱.۹ اهداف فصل

] [

### ۲.۹ پرسپترون

] [

### ۳.۹ آموزش پرسپترون

] [

### ۴.۹ پرسپترون چندلایه

] [

### ۵.۹ عمومیت

] [

## ١٠ فصل

### یادگیری ناظارت‌نشده

١.١٥ مقدمه

] [

٢.١٥ الگوریتم PCA

] [

٣.١٥ الگوریتم t-SNE

] [

٤.١٥ خوشبندی با k-means

] [

٥.١٥ خوشبندی با k-modes

] [

٦.١٥ خوشبندی با k-prototype

] [

## فصل ۱۱

### پروژه دوم

#### ۱.۱۱ اهداف فصل

] [

#### ۲.۱۱ تعبیهی متن

] [

#### ۳.۱۱ فاصله‌ی ویرایش

] [

#### ۴.۱۱ معیار شبه است RBO

] [

## ۱۲ فصل

### بیشتر بدانید

#### ۱.۱۲ نمونه کاهی با NearMiss

] [

#### ۲.۱۲ نمونه افزایی با SMOTE

] [

#### ۳.۱۲ درخت رگرسیون

] [

## واژه‌نامه

( ) .

# **منابع**

( ) . / / /