

موضوع درس: مباحث ویریه

استاد: محمد احمدزاده

دانشجو: ماهکده خاچی - مریم لریستانی نژاد

ترم بهمن ۱۴۰۳

۱۴۰۳/۱۲/۲۰

بخش ۱ Machine Learning  
درس: مباحث ویژه

## A: unsupervised Learning, supervised Learning

چه تفاوتی دارند؟ تفاوت اصلی بین یادگیری نظارت شده و یادگیری بدون نظارت تمایز اصلی بین این دو رویکرد یادگیری ماشین استفاده از مجموعه داده‌های برچسب‌دار است. برای سامره یادگیری نظارت شده supervised Learning از داده‌های ورودی و خروجی برچسب‌دار استفاده می‌کند در حالی که یادگیری بدون نظارت unsupervised Learning این کار را نمی‌کند. در یادگیری نظارت شده هدف پس بینی نتایج برای داده‌های جدید است و تنها از قبل از نوع نتایج برای داده که باید انتظار داشته باشد اطلاع داریم اما در یک الگوریتم یادگیری بدون نظارت هدف این است که از حجم زیادی از داده‌های جدید دست بیاریم الگوریتم خودش تعیین می‌کند که چه چیزی در مجموعه داده مورد نظر متفاوت است یا جالب است و چه الگوهای در آن وجود دارد که باین صورت وجه قرار گیرد.

## B: چرا Feature scaling در الگوریتم‌های Machine Learning

ضروری است؟ مقیاس بندی ویژگی Feature scaling یک تکنیک برای استاندارد سازی ویژگی‌های مستقل Independent Feature موجود در داده‌ها به یک بازه مشخص در یادگیری ماشین Machine Learning است. مقیاس بندی ویژگی‌ها یک گام اساسی در پرسش برداشتن داده‌ها است که تعیین می‌کند تمام ویژگی‌ها در یک مقیاس و بازه مناسب قرار می‌گیرند این فرایند اهمیت زیادی برای بهبود سرعت مدل‌های یادگیری ماشین دارد.

## C: Normalization و standardization چه تفاوتی دارند؟

خرمال سازی و استاندارد سازی هر دو تکنیک‌های پرسش پردازش داده هستند که برای تغییر مقیاس مقادیر ویژگی‌ها در مجموعه داده‌ها استفاده می‌شوند هدف اصلی این تکنیک‌ها تبدیل مقادیر به یک بازه مشخص یا توزیع استاندارد است تا عملکرد الگوریتم‌های یادگیری ماشین بهبود یابد با این حال تفاوت‌های کلیدی بین روش‌ها وجود دارند. خرمال سازی هدف = مقیاس بندی مقادیر ویژگی‌ها به یک بازه مشخص معمولاً بین ۰ و ۱۰۰ فرمول:

$$X_{\text{normalized}} = \frac{(x - \min)}{(x_{\text{max}} - x_{\text{min}})}$$



# بخش 1 Machine Learning : ۱۴۰۳/۱۲/۲۵

Subject:

Date:

درس: مباحث ویژه

## D: چرا Min-Max Normalization برای

مقیاس بندی داده استفاده می شود؟ روش نرمال سازی این روش مناسب داده هایی است که توزیع نرمال ندارند و نیاز به مقیاس بندی در بازه خاص دارند. این روش داده ها را به بازه ای مشخص معمولاً بین ۰ و ۱ تغییر می دهد.

بعضی عملیات در الگوریتم ها: بسیاری از الگوریتم های یادگیری ماشین به ویژه آن هایی که به فاصله ی بین داده ها حساس هستند ویژگی هایی که با مقادیر بزرگ تر می توانند ویژگی های با مقادیر کوچک تر غالب شوند سرعت بخیرین به فرآیند یادگیری: با مقیاس بندی داده ها الگوریتم ها به سرعت به حداقل محلی یا بهینه ترین وزن ها همگرا می شوند یک روش ساده و موثر برای مقیاس بندی داده ها است که در بسیاری از موارد مفید است اما محدودیت هایی نیز دارد که باید در نظر گرفته شود. انتخاب روش مقیاس بندی مناسب به نوع داده ها و الگوریتم یادگیری ماشین مورد استفاده بستگی دارد.

## E: Z-score Normalization چیست و چرا کاربرد دارد؟ روشی در آمار و یادگیری ماشین است که داده ها را به مقیاس با میانگین ۰ و انحراف استاندارد ۱ تبدیل می کند. در این تکنیک تمامی مقادیر که خارج از یک محدوده مشخص باشند دور ریخته می شوند. ۱- مقایسه داده ها با واحدهای مختلف - وقتی داده ها با واحدهای اندازه گیری مختلف جمع آوری شده اند Z-score می تواند آن ها را استاندارد کرده و امکان مقایسه و ترکیب آن ها را فراهم کند. ۲- بعضی عملیات در الگوریتم های یادگیری ماشین: بسیاری از الگوریتم های یادگیری ماشین مانند رگرسیون مقایسه داده ها حساس هستند. یک ریزر قدرتمند برای استاندارد سازی داده ها است که در زمینه های مختلف از علم داده یادگیری ماشین آمار و تحلیل مالی کاربرد دارد.



**F: Regularization در الگوریتم های Machine Learning چیست؟**  
یکی از تکنیک های مهم برای جلوگیری از **overfitting** است. **overfitting** زمانی اتفاق می افتد که مدل به داده های یادگیری الگوریتم کلی داده به داده های آموزشی بیش از حد وابسته می شود و در نتیجه عملکرد ضعیفی روی داده های جدید دارد. **Regularization** با اضافه کردن یک جریمه (**Penalty**) به تابع هزینه (**loss function**) مدل را مجبور می کند تا پارامترهای ساده تری را یاد بگیرد.

**G: Underfitting, overfitting چه مسائلی را در Model-building به وجود می آورند؟**  
**Underfitting** و **overfitting** دو مشکل رایج در ساخت مدل های یادگیری ماشین هستند که می توانند عملکرد مدل را به شدت تحت تأثیر قرار دهند. به تفصیل این مشکلات و تأثیرات آن ها توضیح داده می شود. مشکلات - کاهش تعمیم پذیری: مدل نمی تواند الگوهای کلی را از داده ها استخراج کند و فقط به داده های آموزشی خاص وابسته می شود. ۲- حساسیت به نویز: مدل به نویزهای موجود در داده های آموزشی حساس می شود و این نویزها به عنوان الگوهای مهم در نظر می گیرند.

**H: cross-validation چیست؟**  
**cross-validation** در کنار **Train/Test split** استفاده می شود تا از مشکل **overfitting** و انتخاب تعداد داده های **train** و **test** جلوگیری کند. در واقع **cross-validation** به ما کمک می کند تا ارزیابی دقیق تری از عملکرد مدل داشته باشیم. در روش **Train/Test split** داده ها به دو بخش آموزش (**Train**) و آزمون (**Test**) تقسیم می شوند. به داده های آموزش داده می شود و سپس با داده های آزمون ارزیابی می شود. ارزیابی دقیق تر با استفاده از چندین مجموعه داده آموزش و آزمون می توان ارزیابی دقیق تری از عملکرد مدل داشت.



Subject:

بخش 1 Machine Learning ۱۴۰۳/۱۲/۲۵

Date:

درس: مباحث ویژه

۱: Gradient Descent چگونه کار می کند، تعدادی کاهشی

یک الگوریتم بهینه سازی برای پیدا کردن کمینه یک تابع است. در این الگوریتم کار جابجایی نقطه تعادلی روی تابع آغاز می شود و روی جهت منفی از گرادیان تابع حرکت می کند تا به کمینه محلی سراسری برسد. اصل کار در تیرا دیان کاهشی؛  
۱- مقدار درونی اولیه: ابتدا یک نقطه شروع تعادلی یا بر اساس یک روش خاص برای چارامترهای مدل مانند وزن ها و بایاس ها در یک شبکه عصبی انتخاب می کنیم این نقطه حدس اولیه ما برای بهینه ترین مقدار است. ۲- بهینه سازی محلی اگر ا دیان کاهشی ممکن است در یک بهینه سازی محلی گیر کند و نتواند به بهینه سازی سراسری برسد.

۲: چرا Deep Learning برای پیچیده ترین مسائل استفاده می شود؟

یادگیری عمیق به دلیل مختلف برای مسائل پیچیده استفاده می شود. قابلیت یادگیری ویژگی های پیچیده یا یادگیری عمیق با استفاده از شبکه های عصبی عمیق قادر است به صورت خودکار ویژگی های پیچیده و غیر خطی را از داده ها یاد بگیرد. در روش های عمیق یادگیری ما بین اغلب نیاز به استخراج دستی ویژگی ها توسط انسان بوده است که این کار زمان بر و نیازمند تخصص جدی دارد. یادگیری عمیق به یک ابزار قدرتمند برای حل مسائل پیچیده تبدیل شده است و به طور گسترده در صنایع مختلف مورد استفاده قرار می گیرد.



# بخش ۲: Python Programming ۱۴۰۳/۱۲/۲۰

## درس: مباحث ویژه

**A: چرنا** Python زبان برنامه نویسی محبوب علم داده است. زبان پایتون یک زبان سطح بالا است که در سال ۱۹۹۱ ادای جاگیز بونی زبان ABC است. امروزه در بسیاری از حوزه های برنامه نویسی و روبر کرده و بسیاری از منابع از آن استفاده می کنند. دلیل اصلی محبوبیت آن سادگی یادگیری و بهر دور از پیچیدگی است. به دلیل انبوهی از کتابخانه ها و همچنین داشتن باز بودن خود در یکی از قوی ترین زبان های موجود است. از دیگر امتیازات سادگی این زبان من بودن با آسان کار کردن و همگی هنگام کار با آن اشاره کردیم. این دلیل محبوب دارد که به صورت گروه شروع به فعالیت با زبان برنامه نویسی پایتون می کنند و به سادگی می توانند پروژه را پیش ببرند و کار راحت تر نسبت به نرم افزار های مشابه در پیش دارند.

**B: Numpy و Pandas** چه تفاوتی دارند؟ کتابخانه Pandas یک کتابخانه متن باز Open Source به زبان برنامه نویسی پایتون است که کارایی بالایی در مراحل تحلیل های آماری، تغییر پس پردازش و تقویر سازی داده ها دارد. از طرفی کتابخانه Numpy یک کتابخانه مهم دیگر در پایتون به خصوص برای تحلیل گر های داده Data Scientist است. آرایه Numpy نوعی ساختار داده چند بعدی در پایتون است که می تواند آرایه های یک بعدی از انواع داده ها را مشابه برانده خیره کند. در کتابخانه Pandas ساختار داده ای دو بعدی جدولی و قابل تغییر در پایتون است که می تواند داده های مختلف را ذخیره کند.

**C: چرنا** Matplotlib برای تجسم داده ها استفاده می شود. یک کتابخانه جامع گرافیکی برای ایجاد تجسم های ثابت انیمیشن و تعاملی در پایتون است. کاربرد اصلی این کتابخانه تولید طرح و نمودار از داده ها است. این ویژگی آن را تبدیل به یک ابزار مهم برای تجزیه و تحلیل و تجسم داده ها کرده است. با کمک آن می توان به ماهیت و کیفیت داده ها پی برد. یک کتابخانه جامع گرافیکی برای ایجاد تجسم های ثابت انیمیشن و تعاملی در پایتون است. کاربرد اصلی این کتابخانه تولید طرح و نمودار از داده ها است.



Subject:  
Date:

بخش 2 : Python Programming ۱۴۰۳ / ۱۲ / ۲۰

درس ۴ مباحث ویژه

D: seaborn چرا برای تجسم داده‌های پیچیده کاربرد دارد؟

کتابخانه seaborn به دلایل مختلفی برای تجسم داده‌های پیچیده کاربرد دارد:

- ۱- رابط سطح بالا: seaborn یک رابط سطح بالا برای ایجاد نمودارهای جذاب و آموزنده ارائه می‌دهد. این رابط کاربر را برای کار بر روی آسان‌تر می‌کند تا نمودارهای پیچیده را با حداقل کد ایجاد کند.
- ۲- تجسم توزیع‌های آماری: seaborn ابزارهای قدرتمندی برای تجسم توزیع‌های آماری ارائه می‌دهد. از جمله هسته‌گرام‌ها، نمودارهای تراکم و نمودارهای جعبه‌ای. این ابزارها به کاربران کمک می‌کند تا الگوها و روابط موجود در داده‌ها را شناسایی کنند.

E: چگونه می‌توانید یک Function در Python تعریف کنید؟

هر تابع Function یک وظیفه مشخص دارد که در زمان فراخوانی آن را انجام می‌دهد. مثلاً از تابع len برای شمرش تعداد اشیاء یک لیست استفاده می‌شود. توابع کمک می‌کنند تا کد خود را به بخش‌های کوچک‌تر و قابل مدیریت‌تر تقسیم کند. از تکرار کد جلوگیری کنند و خوانایی کد را افزایش دهند. برای تعریف یک تابع در پایتون از کلمه کلیدی def استفاده می‌شود. در پایتون یک تابع Function یک بلوک کد سازمان یافته و قابل استفاده مجدد است که برای انجام یک وظیفه خاص طراحی شده است. توابع به شما کمک می‌کنند تا کد خود را به بخش‌های کوچک‌تر و قابل مدیریت‌تر تقسیم کنید.



## بخش 2 Python Programming ۱۴۰۲/۱۲/۲۰

درس: مباحث ویژه

**F:** چرا List Comprehension در Python استفاده می شود؟  
یک روش ساده در پایتون برای فیلتر کردن یک لیست بر اساس برخی معیارهای خاص است.  
با استفاده از این امکان می توان در چندین خط کدنویسی صرفه جویی کرد.  
برای خواندن و نوشتن لیست Comprehension را می توان مختصر و قابل فهم برای ایجاد  
لیست ها ارائه داد. در این روش اغلب می توان از جایگزین حلقه های for چند خطی  
شماره گذاری لیست Comprehension معمولاً سریعتر از ساخت لیست با استفاده از حلقه  
های for است به خصوص برای عملیات ساده و تکراری قدرتمند در پایتون است که به شما  
امکان می دهد که تمیز تر و سریع تر و خوانا تر بنویسید.

**G:** چگونه می توان یک CSV file را در Python خواند؟ فایل CSV  
به عنوان یک فایل متنی با قالبی پایتون به نام open باز می شود که یک شی فایل  
را برای نگه داشتن برای پردازش به مثابه reader منتقل می شود. این محتوای  
فایل CSV employee-birthday.csv است. فایل CSV معمولاً توسط  
برنامه های ایجاد می شوند که حجم زیادی از داده را مدیریت می کنند. تجزیه فایل  
های CSV با کتابخانه CSV در پایتون این کتابخانه عملکرد را برای خواندن  
و نوشتن در فایل های CSV فراهم می کند.

**H:** JSON و XML چه تفاوتی دارند؟ JSON با XML با هم متفاوت است.  
در JSON یک بسته نداریم. حجم فایل های JSON کمتر است. فایل های JSON  
زمان خواندن و نوشتن کمتری را می گیرند در فایل XML امکان ذخیره سازی  
آرکایه ها نیست. فایل های JSON توسط JavaScript استفاده می شوند.  
XML نسبت به JSON از ساختار دستوری طولانی تر برخوردار است.  
بر خلاف XML فرمت JSON می تواند از آرایه استفاده کند.

✓



# بخش ۲: Python Programming ۱۴۰۳/۱۲/۲۰

## درس، مباحث و تیر

**A: چرا Python زبان برنامه نویسی محبوب علم داده است؟ زبان پایتون یک**  
زبان سطح بالا است که در سال ۹۹۱ ادای جاگیزین زبان ABC است امروزه در بسیاری از حوزه  
های برنامه نویسی و روخته و بسیاری از منابع از آن استفاده می کنند دلیل اصلی محبوبیت آن  
سادگی یادگیری و به روز رسانی پیوسته است به دلیل انبوهی از کتابخانه ها و همچنین متن باز بودن  
خود یکی از قوی ترین زبان های موجود است از دیگر امتیازات سادگی این زبان من تون با آسانی  
کارگر و هی هنام کار با آن آسان تر می شود به این دلیل محبوب دارد که به صورت گروه شروع به فعالیت  
خازبان برنامه نویسی پایتون می کنند و سادگی می تواند پروژه را پیش ببرند و کار راحت تر شد به نرم افزار  
های مشابه در پیش دارند

**B: NumPy و Pandas چه تفاوتی دارند؟ کتابخانه Pandas یک کتابخانه**  
متن باز Open Source به زبان برنامه نویسی پایتون است که کارایی بالا در مراحل تحلیل های  
آماره ای تغییر پیش پردازش و تقویر سازی داده ها دارد از طرفی کتابخانه NumPy یک کتابخانه  
مهم دیگر در پایتون به خصوص برای تحلیل گر های داده Data Scientist است  
آرایه NumPy نوعی ساختار داده چند بعدی دارد که می تواند آرایه های  
از انواع داده ها را مشابه برانده خیره کند در Pandas ساختار داده ای دو بعدی جدوی و  
قابل تغییر در پایتون است که می تواند داده های مختلف را ذخیره کند

**C: چرا Matplotlib برای تجسم داده ها استفاده می شود؟ یک کتابخانه**  
جامع گرافیکی برای ایجاد تجسم های ثابت انیمیشن و تعاملی در پایتون است کاربرد  
اصلی این کتابخانه تولید طرح و نمودار از داده ها است این ویژگی آن را تبدیل به  
یک ابزار مهم برای تجزیه و تحلیل و تجسم داده ها کرده است با کمک آن  
می توان به ماهیت و کیفیت داده ها پی برد یک کتابخانه جامع گرافیکی  
برای ایجاد تجسم های ثابت انیمیشن و تعاملی در پایتون است کاربرد  
اصلی این کتابخانه تولید طرح و نمودار از داده ها است