# in the name of god

علی اکبر مکاریان

مباحث ویژه

کارشناسی پیوسته

مدرس : مهندس محمد احمدزاده

**بخش ۱ ( ماشین لرنینگ ) :**

**Supervised Learning و Unsupervised Learning چه تفاوتی دارند؟**

در یادگیری ماشین دو حالت اصلی وجود دارد با اسم های یادگیری نظارت شده و یادگیری بدون نظارت که در حالت اول داده ها برچسب گذاری میشوند و آموزش میبینن تا ورودی ها و خروجی ها را به هم مرتبط کنند و در یادگیری بدون نظارت مدل داده ها را برچسب گذاری نمیکند و بدنبال ساختار های پنهان در داده ها است مانند الگوریتم خوشه بندی.

چرا Feature Scaling در الگوریتم‌های Machine Learning ضروری است؟

این ویژگی باعث میشود که مقیاس متغییر ها یکسان بشه و از تاثیر بیش از حد ویژگیهایی با مقادیر زیاد جلوگیری کنه

Standardization و Normalization چه تفاوتی دارند؟

در استانداردسازی داده ها رو جوری تغییر میدیم که میانگین انحراف معیارشون یک بشه . استاندارد سازی برای الگوریتم هایی که فرض میکنن داده های توزیع نرمال دارن مثل شبکه های عصبی مناسبه

و در نرمال سازی داده های در یک بازه مشخص هستن معمولا بین ۰ تا ۱ . نرمال سازی برای مقیاس ویژگی های خیلی حساس مناسبه

چرا**Min-Max Normalization** برای مقیاس‌بندی داده‌ها استفاده می‌شود؟

نرمال سازی Min max مقادیر ویژگی ها را در یک بازه مشخص که معمولا 0 و ۱ هست مقیاس میکند. این روش برای الگوریتم هایی که به مقیاس ویژگی ها حساس هستن مثله شبکه های عصبی مفید است با این حال به روش مقادیر پرت حساس است و ممکن تحت تاثیرش قرار بگیره



**Z-Score Normalization** چیست و چرا کاربرد دارد؟

در این روش داده را جوری تغییر میدهیم که میانگین آن ها صفر و انحراف معیارشان یک شود. این روش برای الگوریتم هایی که فرض میکنند داده ها توزیع نرمال دارند مانند رگرسیون خطی مناسب است.



**Regularization** در الگوریتم‌های **Machine Learning** چیست؟

روشی است که با اضافه کردن یک جریمه به تابع هزینه از overfitting مدل جلوگیری میکند و این جریمه معمولا به صورت اضافه کردن نرم های L1 و L2 به تابع هزینه اعمال میشود که به ترتیب به آن ها Lasso و Ridge میگویند. با این کار مدل ساده تر میشود و بهبود میابد.



**Overfitting** و **Underfitting** چه مشکلاتی را در **Model-building** به وجود می‌آورند؟

در overfitting مدل به طور بیش از حد به جزعیات داده های آموزشی میپردازد و الگوهای نویزی را نیز یاد میگیرد که منجر به عملکرد ضعیف روی داده های جدید میشود.

در underfitting مدل به طور کافی الگوهای موجود در داده های آموزشی را یاد نمیگیرند و بنابرین نمیتوانند به خوبی روی داده های آموزشی و جدید عمل کنند.



**Cross-Validation** چرا در **Train/Test Split** کاربرد دارد؟

روشی است که در آن داده ها به چند قسمت تقسیم میشوند و مدل بصورت مکرر روی ترکیبات مختلفی از این بخش ها آموزش داده شده و ارزیابی میشود. این روش به ارزیابی دقیق تر عملکرد مدل و کاهش وابستگی به یک تقسیم بندی خاص از داده ها کمک میکند.



**Gradient Descent** چگونه کار می‌کند؟

این یک نوع الگوریتم بهینه سازی است که هدف آن کمینه کردن تابع هزینه با بروزرسانی مکرر پارامتر های مدل در جهت مخالف گرادیان تابع هزینه است این فرایند تا زمانی ادامه میابد که تابع هزینه به مقدار کمینه محلی یا جهانی خود برسد



چرا **Deep Learning** برای پیچیده‌ترین مسائل استفاده می‌شود؟

یادگیری عمیق با استفاده از شبکه های عصبی عمیق قادر است ویژگی های پیچیده را از داده ها استخراج کند و این قابلیت به آن اجازه میدهد تا در مساعلی مانند : مدل سازی روابط پیچیده – استخراج ویژگی های خودکار – مقیاس پذیری داده های بزرگ و . . . و در بهداشت و درمان و خودروسازی و مساعل بسیار زیادی باعث پیشرفته خیلی زیادی شده است.



**بخش ۲ ( پایتون ) :**

چرا پایتون زبان برنامه نویسی محبوب علم داده است؟

پایتون به دلیل دارا بودن کتابخانه های خیلی زیاد در بحث علم داده مثل نام پای و سایکیت و پانداس و . . . و همچنین کامیونیتی فعال و سادگی و خوانایی بالا بسیار محبوب است.



پانداس و نام پای چه تفاوفتی دارند ؟

هر دو کتابخانه های مربوط به علم داده هستن اما در ساختار داده و نوع داده و عملیات فرق میکنند مثلا در نام پای آرایه های چند بعدی وجود دارد و پانداس از ساختار هایی مثل دیتافریم را به ما میدهد و پانداس توابع بیشتری برای فیلترکردن و گروهبندی و تجزیه تحلیل برای دیتاها را دارد



چرا matplotlib برای تجسم داده ها استفاده میشود ؟   
  
چون این کتابخونه استاندارد برای تجسم داده ها است و دارای انعطاف پذیری بالا است مثلا میتواند نمودارهای خطی و میله و پراکنده را فراهم کند و قابلیت تنظیم جزعیات مختلف برای نمودارها از جمله رنگ و اندازه و . . . را دارا است و به خوبی با کتابخانه های نام پای و پانداس هماهنگ میشود.



چرا List Comprehension در پایتون استفاده میشود؟

یک روش مختصر و کارآمد برای ایجاد لیست ها در پایتون است که کدنویسی را سریع تر و خواناتر میکند و بجای استفاده از حلقه های تکرار برای افزودن عناصر به لیست میتوان استفاده کرد.



چگونه میتوان یک فایل CSV را در پایتون خواند؟  
  
با استفاده از کتابخانه pandas و یا csv که در پایتون وجود دارد میتوان اینکار را کرد.



تفاوت جیسان و ایکس ام ال ؟  
  
هر دو فرمت های تبادل داده هستند اما ساختارهای متفاوتی دارند مثلا در json ما ساختار داده ای مانند شی و آرایه را داریم و در xml بر اسا تگ های درختی است و همچنین json خوانایی خیلی بالاتری نسبت به xml دارد

جیسان بیشتر در وب و تبادل داده های سبک استفاده میشود و xml ساختار پیچیده تر و دقیق تری دارد و نیازمند ساختار دهی است.

منابع :

Faradars Blog

7 Learn

Mongard

DATAYAD

WikiPedia

Danup

Bigdataworld

Hamroyesh

Amir.today