

بسمه تعالی

سپیده قاضیانی

تمرین های فصل اول

پاسخ سوال یک)

ماشین لرزینگ یکی از زیر مجموعه های هوش مصنوعی می باشد که سیستم ها به کمک آن می توانند بطور اتوماتیک و بدون نیاز به هیچگونه برنامه نویسی یادگیری و پیشرفت کنند . تمرکز اصلی یادگیری ماشین بر توسعه برنامه های کامپیوتری می باشد که بتوانند به اطلاعات دسترسی یافته و از آن برای یادگیری خود بهره ببرند .

از آنجایی که استفاده از نیروی انسانی هزینه بر می باشد و دقت نیروی انسانی در یک فرآیند طولانی مدت پایین می آید از ماشین لرزینگ استفاده می شود زیرا تکرار پذیری بالاتری دارد طوری که یک کار را هزار بار با یک تکرار و یک دقت انجام میدهد .

مزایای ماشین لرزینگ:

- ۱- به راحتی روندها و الگوها را شناسایی میکند
- ۲- هیچ مداخله انسانی نیاز نیست (اتوماسیون)
- ۳- بهبود مستمر دقت و کارایی درحین کار و کسب تجربه
- ۴- مدیریت داده های چند بعدی و چند متغیره
- ۵- کاربردهای گسترده

معایب ماشین لرزینگ :

- نیاز به مجموعه داده های گسترده
- ۲- نیاز به زمان و منابع کافی جهت یادگیری و پیشرفت
- ۳- نیاز به دقت در تفسیر نتایج
- ۴- حساسیت به خطای بالا

تفاوت:

در برنامه نویسی سنتی برنامه نویس برنامه را به طور کامل به کامپیوتر میداد و میگفت چکار کند در واقع فرآیند کاملاً دستی بود و برنامه نویس برنامه را خلق میکرد
اما در یادگیری ماشین فرآیند کاملاً اتوماتیک می باشد که ما دیگر به کامپیوتر برنامه ای نمی دهیم و در عوض ورودی و خروجی را میدهیم و برنامه را میگیریم .

پاسخ سوال دو)

بیش برآزش یکی از خطاهای مدل سازی در علم داده است. این خطا هنگامی اتفاق می افتد که مدل ویژگی های داده های آموزشی را به جای یادگیری حفظ کرده باشد یعنی بیش از حد روی آن آموزش دیده باشد در نتیجه این مدل فقط در مجموعه ی داده های آموزشی مفید خواهد بود و نه در مجموعه های داده های دیگر که هنوز آن ها را ندیده است.

مدل یادگیری ماشین می تواند داده های جدید را پیش بینی کند که به آن تعمیم گفته می شود. هدف کلی این است که مدل یک گروه از تصاویر را ببیند و آن را به خوبی تعمیم دهد تا بتواند تصاویر جدید را پیش بینی کند.

دلایل ایجاد بیش برآزش :

۱-مدل بیش از حد پیچیده است و ویژگی های هم خط را در بر میگیرد که واریانس داده های ما را افزایش می دهد.

۲-تعداد ویژگی های داده های ما بیشتر یا برابر با تعداد داده است

۳-حجم داده بسیار کم است

۴-داده بیش برآزش نشده تمیز نیست و نویز دارد.

راه های مقابله با بیش برآزش:

۱-نگه داشتن داده ها: بجای استفاده از همه ی داده های خود مجموعه ی داده ها را به دو قسمت تقسیم میکنیم (آموزش و تست)

۸۰ درصد برای آموزش و ۲۰ درصد برای تست است مدل خود را آموزش میدهم تا زمانی که تا زمانی که عملکرد خوبی داشته باشیم این امر نشان دهنده قابلیت تعمیم خوب مدل می باشد زیرا مجموعه تست نشان دهنده ی داده های دیده نشده ای است که برای آموزش استفاده نشده اند که این مجموعه به داده ای نسبتاً بزرگ نیاز دارد.

۲-اعتبار سنجی متقابل داده ها : می توانیم مجموعه داده مان را به تقسیم کنیم در این روش اجازه می دهیم یکی از گروه های مجموعه تست و گروه دیگر به عنوان مجموعه آموزشی باشد این روند را تا وقتی که هر گروه جداگانه یک بار به عنوان مجموعه ی تست استفاده شود تکرار می کنیم .

۳-افزایش داده ها : یک مجموعه داده بزرگ تر از بیش برآزش جلوگیری می کند اگر نتوانیم داده های بزرگتری جمع کنیم و به داده هایی که در مجموعه ی فعلی خود داریم محدود باشیم می توانیم برای افزایش اندازه داده مان از روش افزایش داده استفاده کنیم .

۴-انتخاب ویژگی : اگر فقط تعداد محدودی نمونه آموزشی داشته باشیم که هر یک از آن ها تعداد زیادی ویژگی داشته باشند فقط باید مهم ترین ویژگی ها را برای آموزش انتخاب کنیم تا مدل ما به یادگیری تمامی ویژگی ها مجبور نباشد که سرانجام بیش

برازش اتفاق بیفتد که می توانیم به سادگی ویژگی های مختلف را آزمایش کنیم مدل های جداگانه ای را برای این ویژگی ها آموزش دهیم و قابلیت تعمیم مدل ها را ارزیابی کنیم یا از یکی از روش های مختلف انتخاب ویژگی استفاده کنیم .

۵- **نظم دهی** : نظم دهی تکنیکی است که شبکه ی ما را در یادگیری مدلی که بیش از حد پیچیده است و ممکن است به بیش برازش بینجامد محدود می کند درنظم دهی می توانیم برای تابع زبان یک مجازات در نظر بگیریم تا ضرایب برآورده شده را به سمت صفر سوق دهیم

۶- **حذف لایه ها در مدل** : یک مدل بیش از حد پیچیده به احتمال زیاد به بیش برازش خواهدانجامید بنابراین ما می توانیم با حذف لایه ها بطور مستقیم از پیچیدگی مدل بکاهیم و اندازه مدل خود را کاهش دهیم .

۷- **دراپ اوت** : این تکنیک نوعی نظم دهی محسوب می شود در برخی از لایه ها برخی نودها را در نظر نمی گیریم با استفاده از دراپ اوت می توانیم یادگیری متقابل میان نودها را که ممکن است به بیش برازش بینجامد کاهش دهیم با وجود این برای اجرای این تکنیک به ایپاک های بیشتری برای هم گرا شدن مدل خود نیاز داریم .

۸: **توقف زود هنگام** :

راه های جلوگیری از بیش برازش :

۱- **کاهش تنظیم** : چندین روش مختلف مانند l1 و تنظیم lasso دراپ اوت و غیره وجود دارد که به کاهش نویز در مدل کمک می کند با این حال اگر ویژگی های داده بیش از حد یکنواخت شوند مدل قادر به شناسایی روند غالب داده نیست و این موضوع به کم برازش می انجامد با کاهش میزان تنظیم پیچیدگی و تنوع مدل بیشتر می شود و امکان آموزش موفقیت آمیز فراهم می شود .

۲ **افزایش مدت زمان آموزش مدل** : توقف زود هنگام آموزش می تواند به مدل کم برازش بینجامد بنابراین با افزایش مدت زمان آموزش می توان از بروز این مشکل جلوگیری کرد آموزش بیش از حد مدل به مشکل بیش برازش می انجامد پس باید تعادلی میان این دو ایجاد کنیم .

۳ **انتخاب ویژگی** : در هر مدلی ویژگی های خاصی برای تعیین نتیجه ی مشخص استفاده می شود اگر ویژگی های کافی وجود نداشته باشد باید ویژگی های بیشتر یا ویژگی هایی با اهمیت بیشتر به مدل اضافه شوند .

پاسخ سوال سه)

Polynomial features , power transformations , and binning .

پاسخ سوال چهار)

Voting and averaging

Bagging

Boosting

Stacking