

به نام خدا

نام درس:
جبر خطی

نام استاد:
جناب دکتر حمید رسولی

ارائه دهنده:
شروین ایران عقیده

زمان برگزاری:
پنج شنبه (۱۳:۰۰ تا ۱۵:۴۵)

مقدمه

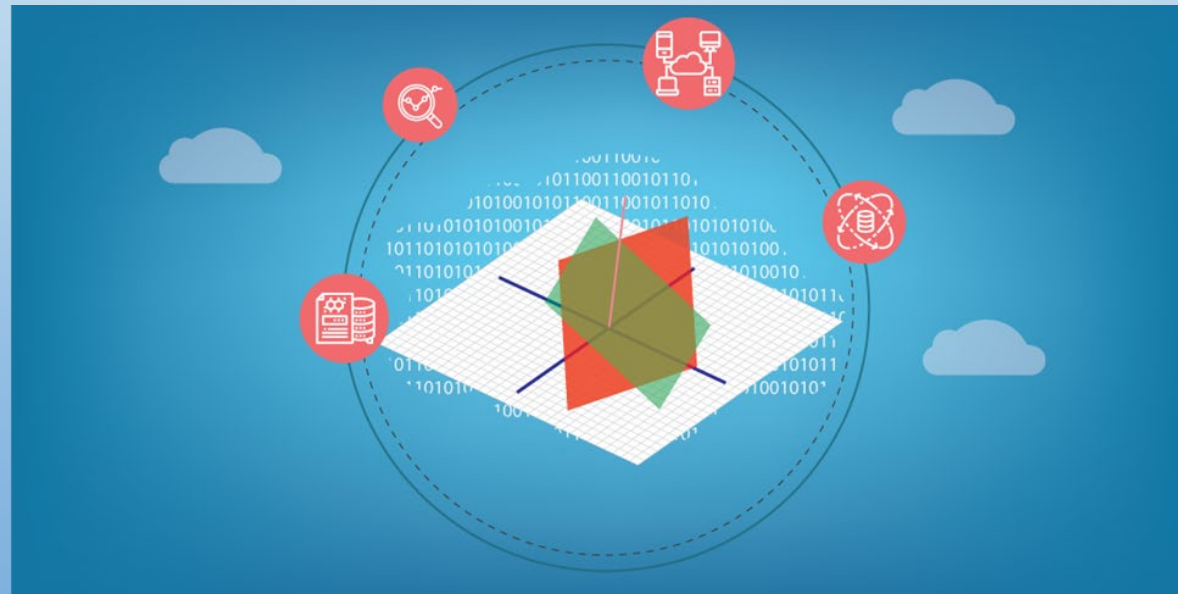
جبر خطی حوزه‌های اصلی علم داده از جمله زمینه‌های داغ پردازش زبان طبیعی و بینایی رایانه را تقویت می‌کند اما اغلب نادیده گرفته می‌شود.

وقتی زبان‌های برنامه‌نویسی برای علم داده بسته‌های زیادی را برای کار با داده‌ها ارائه می‌دهند، مردم زیاد به جبر خطی توجه نمی‌کنند که این یک اشتباه است.

جبر خطی پشت همه الگوریتم‌های قدرتمند یادگیری ماشینی می‌باشد.

جبر خطی یک چرخ دنده حیاتی در مجموعه مهارت‌های دانشمندان داده است.

جبر خطی یکی از بلوک‌های اساسی علم داده می‌باشد و در بسیاری از زمینه‌ها کاربرد دارد.



جبر خطی در یادگیری ماشین

در اینجا این سوال مطرح می‌شود که جبر خطی در یادگیری ماشین کجا قرار می‌گیرد؟

در ادامه به کاربردهای جبر خطی در یادگیری ماشین پرداخته می‌شود.

• تابع ضرر (Loss Function)

لازم است کاملاً با نحوه مطابقت یک مدل، مثلاً یک مدل رگرسیون خطی، با داده‌های داده شده آشنا بود:

شما با یک تابع پیش‌بینی دلخواه (یک تابع خطی برای مدل رگرسیون خطی) شروع می‌کنید.

از آن بر روی ویژگی‌های مستقل داده‌ها برای پیش‌بینی خروجی استفاده کنید.

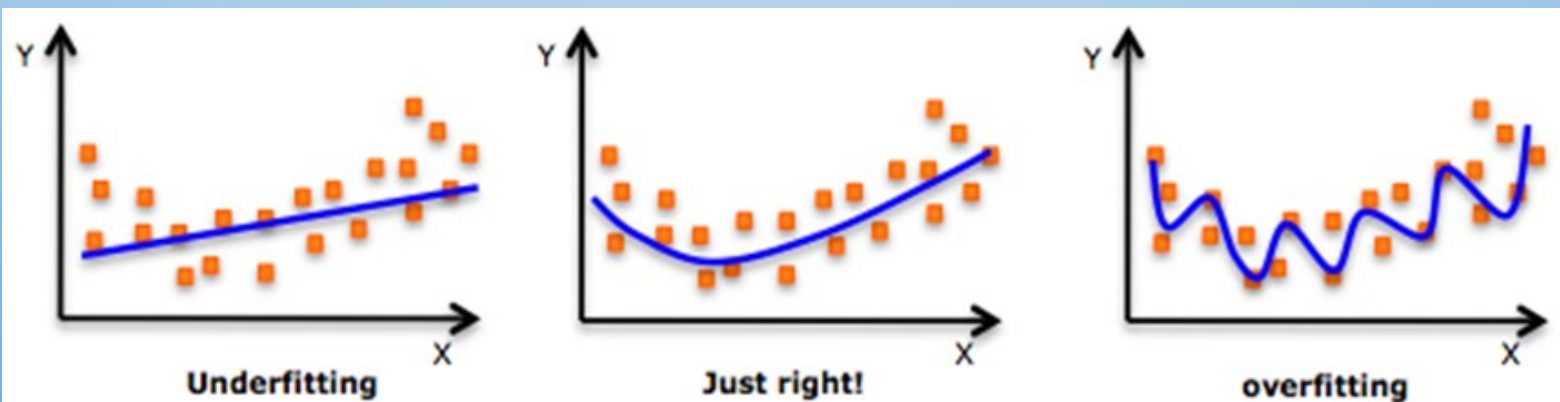
محاسبه کنید که خروجی پیش‌بینی شده چقدر با خروجی واقعی فاصله دارد.

یک تابع ضرر، یک برنامه کاربردی از مقدار عددی بردار در جبر خطی است. مقدار عددی بردار به سادگی می‌تواند بزرگی آن باشد.

جبر خطی در یادگیری ماشین

منظم سازی (Regularization)

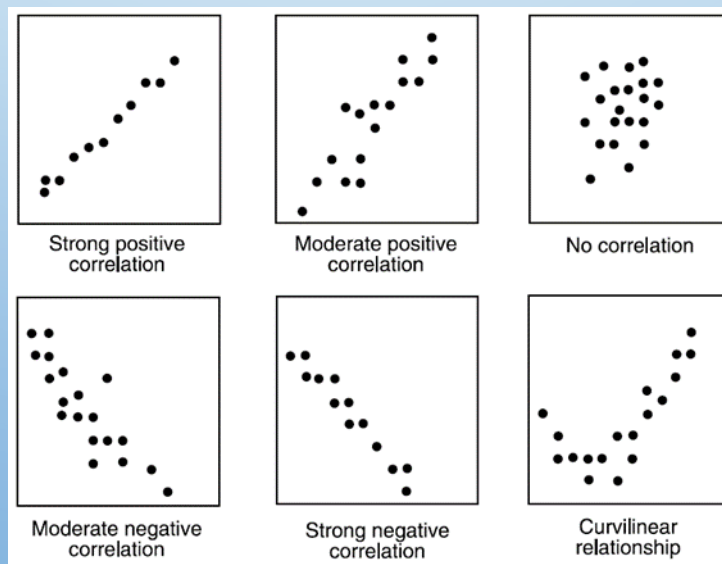
- Regularization یک مفهوم بسیار مهم در علم داده است.
- این تکنیکی است که برای جلوگیری از برازش بیش از حد مدل ها استفاده می شود.
- Regularization در واقع یکی دیگر از کاربردهای مقدار عددی است.
- زمانی گفته می شود که یک مدل بیش از حد برازش می کند که به خوبی با داده های آموزشی مطابقت داشته باشد.
- چنین مدلی با داده های جدید عملکرد خوبی ندارد زیرا حتی نویز در داده های آموزشی را نیز یاد گرفته است.
- نمی تواند داده هایی را که قبلاً ندیده است تعمیم دهد
- تصویر زیر این ایده را به خوبی خلاصه می کند:



جبر خطی در یادگیری ماشین

ماتریس کوواریانس

- تجزیه و تحلیل دو متغیره گام مهمی در اکتشاف داده است.
- کوواریانس یا همبستگی معیارهایی هستند که برای مطالعه روابط بین دو متغیر پیوسته استفاده می‌شوند.
- کوواریانس جهت رابطه خطی بین متغیرها را نشان می‌دهد.
- کوواریانس مثبت نشان می‌دهد که افزایش یا کاهش در یک متغیر در متغیر دیگر با همین امر همراه است.
- کوواریانس منفی نشان می‌دهد که افزایش یا کاهش در یکی با عکس دیگر در دیگری همراه است.



جبر خطی در یادگیری ماشین

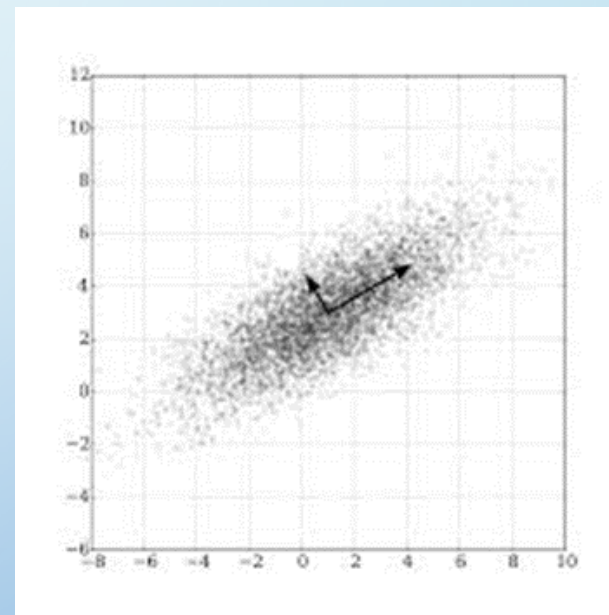
دسته‌بندی ماشین بردار پشتیبان

- یکی از رایج‌ترین الگوریتم‌های طبقه‌بندی است که به طور منظم نتایج چشمگیری ایجاد می‌کند.
- این کاربرد مفهوم فضاها را برداری در جبر خطی است.
- ماشین بردار پشتیبان یا SVM یک طبقه‌بندی متمایز است که با یافتن سطح تصمیم کار می‌کند.
- این یک الگوریتم یادگیری ماشینی تحت نظارت است.
- در این الگوریتم، هر مورد داده را به صورت نقطه‌ای در فضای n بعدی رسم می‌کنیم که مقدار هر ویژگی مقدار یک مختصات خاص باشد.

جبر خطی در یادگیری ماشین

تجزیه و تحلیل مؤلفه اصلی (PCA)

- تجزیه و تحلیل اجزای اصلی یا PCA ، یک تکنیک کاهش ابعاد بدون نظارت است.
- PCA جهت حداکثر واریانس را پیدا می کند و داده ها را در امتداد آن ها برای کاهش ابعاد طرح می کند.
- بدون پرداختن به ریاضیات، این جهت ها بردارهای ویژه ماتریس کوواریانس داده ها هستند.
- بردارهای ویژه برای یک ماتریس مربعی، بردارهای غیر صفر خاصی هستند که جهت آن ها حتی پس از اعمال تبدیل خطی (به معنی ضرب) با ماتریس تغییر نمی کند.



جبر خطی در یادگیری ماشین

پردازش زبان طبیعی (NLP)

- پردازش زبان طبیعی (NLP) زیرشاخه‌ای از زبان‌شناسی، علوم رایانه و هوش مصنوعی است که به تعامل بین رایانه‌ها و زبان انسان، به ویژه نحوه برنامه‌ریزی رایانه‌ها برای پردازش و تجزیه و تحلیل مقادیر زیادی از داده‌های زبان طبیعی مربوط می‌شود.
- هدف، رایانه‌ای است که قادر به "درک" محتویات اسناد، از جمله تفاوت‌های ظریف متنی زبان درون آن‌ها باشد.
- سپس این فناوری می‌تواند اطلاعات و بینش‌های موجود در اسناد را به‌طور دقیق استخراج کند و خود اسناد را دسته‌بندی و سازمان‌دهی کند.
- چالش‌های پردازش زبان طبیعی اغلب شامل تشخیص گفتار، درک زبان طبیعی و تولید زبان طبیعی است.

جبر خطی در یادگیری ماشین

تعبیه سازی کلمات

- الگوریتم های یادگیری ماشینی نمی توانند با داده های متنی خام کار کنند.
- برای ایجاد ورودی های مدل، باید متن را به برخی ویژگی های عددی و آماری تبدیل کنیم.
- راه های زیادی برای مهندسی ویژگی های داده های متنی وجود دارد، مانند:
- فراویژگی های یک متن، مانند تعداد کلمات، تعداد کاراکترهای خاص و غیره .
- ویژگی های NLP متن با استفاده از برچسب های قسمت های گفتار و روابط گرامر مانند تعداد اسم های خاص.
- نمادهای برداری کلمه یا تعبیه سازی کلمات.
- تعبیه سازی کلمات راهی برای نمایش کلمات به عنوان بردارهای کم بعدی اعداد در حالی که محتوای آنها در سند حفظ می شود.

درک چگونگی کاربرد جبر خطی در هوش مصنوعی

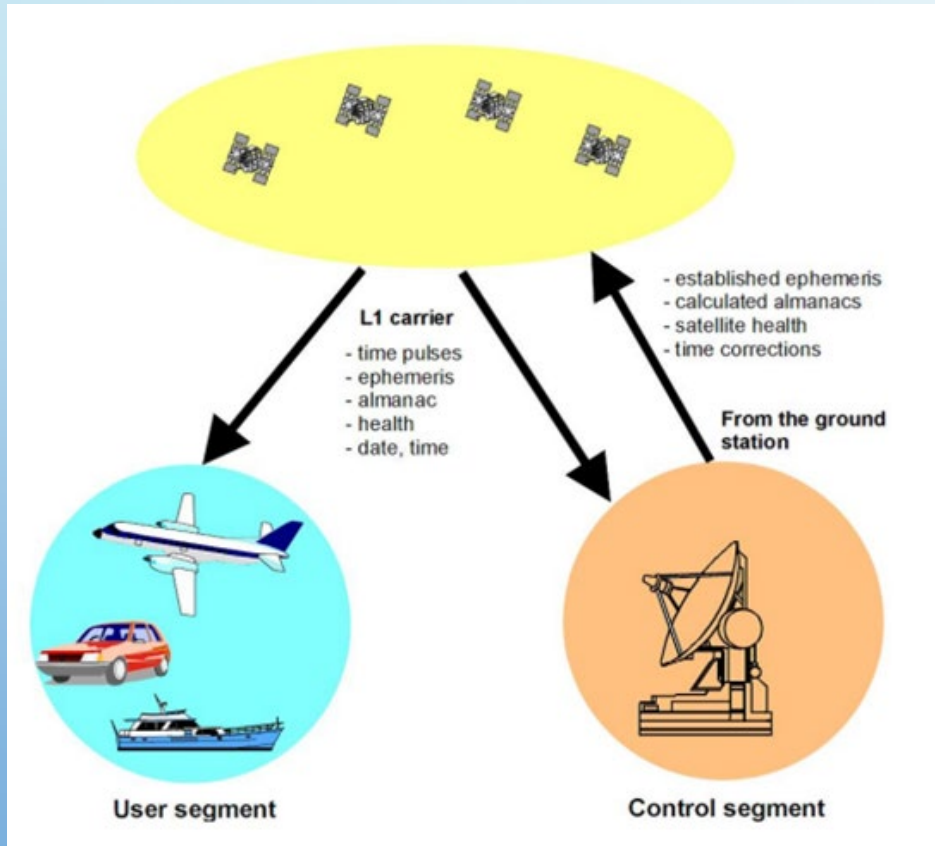
جبر خطی یا اشیای ریاضی، بردارها، ماتریس‌ها و تنسورها هستند.

بسته به ابعاد داده‌ها باید شی مناسب را برای ذخیره و پردازش انتخاب نمود.

قبل از شروع نحوه استفاده از اشیای ریاضی در هوش مصنوعی، بهتر است جبر خطی را بازخوانی نمود.

- **نمایش داده‌ها:** بر حسب بردار اجسام ریاضی، ماتریس و تنسور توضیح داده شده است.
- **مجموعه داده:** مجموعه‌ای از نمونه‌ها یا نقاط داده یا اشیا است.
- **ماتریس طراحی:** یک مجموعه داده را می‌توان از طریق ماتریس طراحی توصیف کرد. ماتریس طراحی، ماتریسی است که در هر سطر شامل یک مثال متفاوت است.
- **پردازش داده‌ها:** قبل از اینکه از مجموعه داده‌ها در الگوریتم‌های یادگیری ماشین خود یا در هر زیر زمینه هوش مصنوعی استفاده کنیم، لازم است مجموعه داده آماده (پاک شده و فیلتر شده) باشند.

کاربرد جبر خطی در سیستم موقعیت یاب جهانی یا GPS



GPS یا سیستم موقعیت یاب جهانی یک سیستم ماهواره‌ای است که امکان یافتن موقعیت جغرافیایی، سرعت و برخی از شرایط محیطی را فراهم می‌کند.

با توجه به تکنولوژی‌های ارائه شده در محصولات مختلف به راحتی می‌توان گفت که در دنیای امروز GPS در همه جا وجود دارد.

این تکنولوژی محبوب در اتومبیل، تلفن هوشمند، ساعت و بسیاری از اکسسوری‌ها و گجت‌های هوشمند حرف اول را می‌زند.

یک سیستم GPS از سه قسمت کلی تشکیل شده است:

- بخش فضا (شامل مجموعه ماهواره‌ها)
- بخش کنترل (شامل همه ایستگاه‌های کنترل و مانیتورینگ زمینی)
- بخش کاربری (شامل همه گیرنده‌های نظامی و غیر نظامی)

لازم به ذکر است که محاسبات موقعیت در سیستم GPS با جبر خطی می‌باشد.

دیگر کاربردهای جبر خطی





با تشکر از توجه شما