جلسه 12

PCA(principal component analysis) : در این حالت unsupervised هست و نباید به صورت پیش بینی به مساله نگاه کرد. هدف اصلی پیش بینی نیست ولی میتوان از خروجی برای پیش بینی استفاده کرد.(تحلیل مولفه های اصلی)

در unsupervised ها هدف بررسی ساختار دادهاست و نه لزوما پیش بینی .

یه مبحث وجود داره به نام cursive dimensionality که بیان میکنه بعدهای زیاد باعث میشود ما نتونیم prediction خوب داشته باشیم و حتی در unsupervisedها نیز تحلیل خوبی داشته باشیم.در تئوری تعداد زیاد متغییر خطا هاور بالاست اما لزوما همیشه اینگونه نیست.(high dimensionality)

این الگوریتم به این صورت کار میکنه که یه دیتا ست میگیزه به ابعاد n\*pو یه خروجی برمیگردونه به ابعاد n\*q که خروجی این الگوریتم هستش . به feature selection معروف است.

PCA همانند SVD میتونه کمک کنه که noise reduction رو انجام بدیم.

یه نکته که باید بهش خیلی توجه بشه اینه که اگر دوتا خط بر هم عمود باشن Orthogonal هستن و استقلال دارن به این معنی که درPCA اگر دو محور عمود برهم نتخاب کنیم اگر عمود برهم باشن یه اطلاعاتی که میگیریم خیلی کاملا از هم مستقل هستند.

هر چی واریانس بالاتر باشه اطلاعات بیشتری رو از داده ها کاور میکنه به همین دلیل در PCA تلاش میکنیم که واریانس رو ماکزیمم کنیم.

در فرمول مذکور wها همون راستاهای ما خواهند بود.

تفاوت اسکیل متغییرها در PCA تاثیر گذاراست.