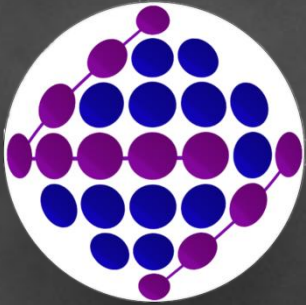


Principal Component Analysis



فصل سیزدهم



شرکت فروش مصنوعي رسا

مدرس : خريد فاشمي نژاد

Principal Component Analysis Theory



جلسہ سی و نہم



شرکت فروش مصنوعي رسا

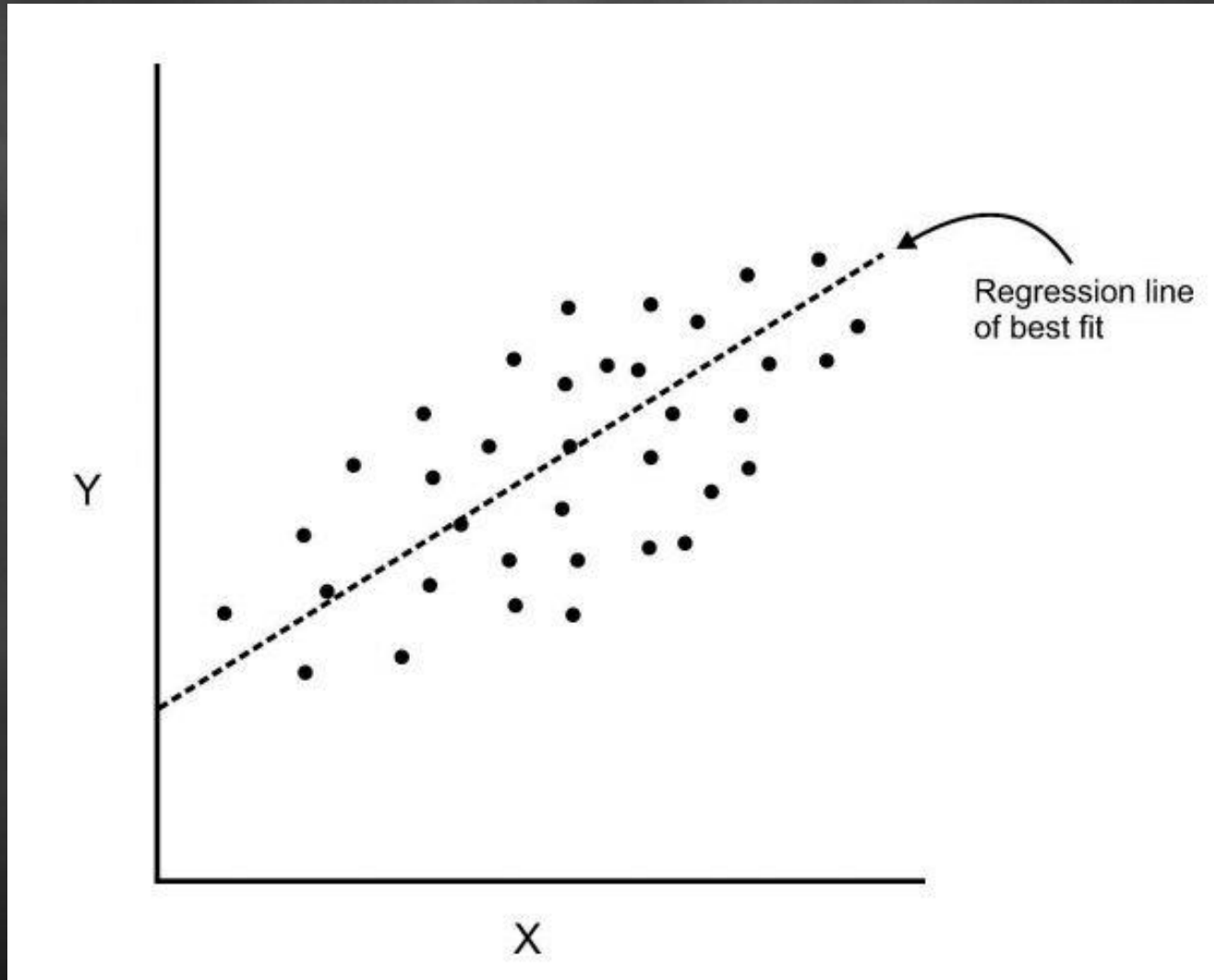
مدرس : فرید فاشمی نژاد

• **پایید ایده اصلی پشت تجزیه و تحلیل مولفه های**

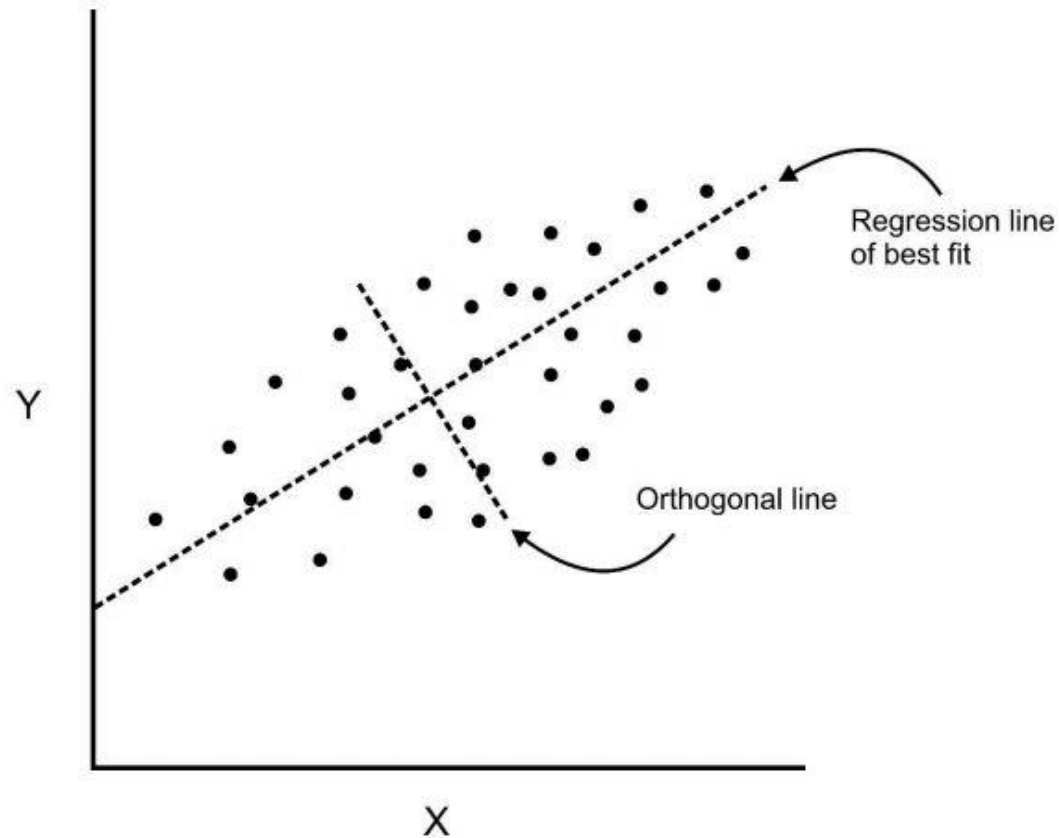
اصلی (PCA) را مورد بحث قرار دهیم. این یک تکنیک آماری بدون نظارت است که برای بررسی روابط متقابل مجموعه ای از متغیرها به منظور شناسایی ساختار اساسی آن متغیرها استفاده می شود. گاهی اوقات به عنوان یک تحلیل عاملی کلی نیز شناخته می شود.

- زمانی رگرسیون خطی بهترین تناسب را برای مجموعه داده \mathcal{D} تعیین می کند که فاکتور تحلیل، چندین خط متعامد با بهترین تناسب را با مجموعه داده \mathcal{D} تعیین کند.
- متعامد به معنی "در زاویه درست" است.
- در واقع خطوط بر یکدیگر در فضای n بعدی عمود هستند.
- به اندازه متغیر \mathcal{D} ابعاد وجود دارد، بنابراین در یک مجموعه داده با d متغیر فضای نمونه d بعدی است.

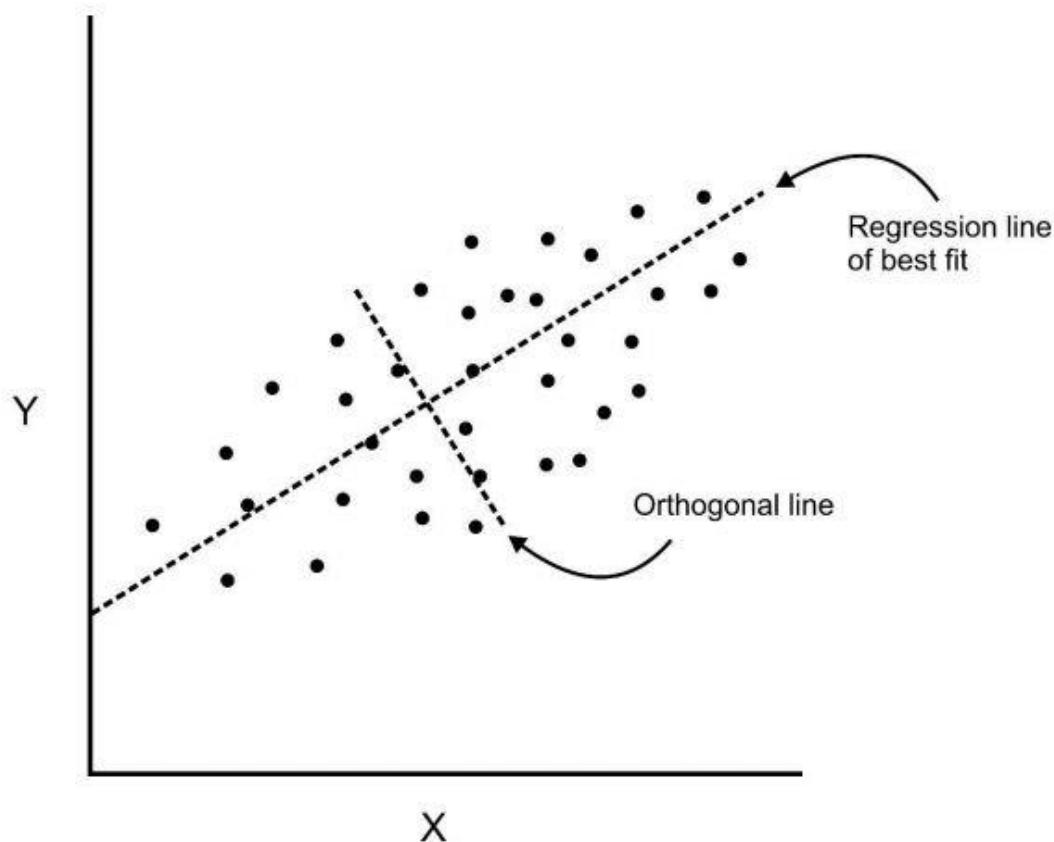
در اینجا ما برخی از داده ها را با دو ویژگی X و Y ترسیم کرده ایم.



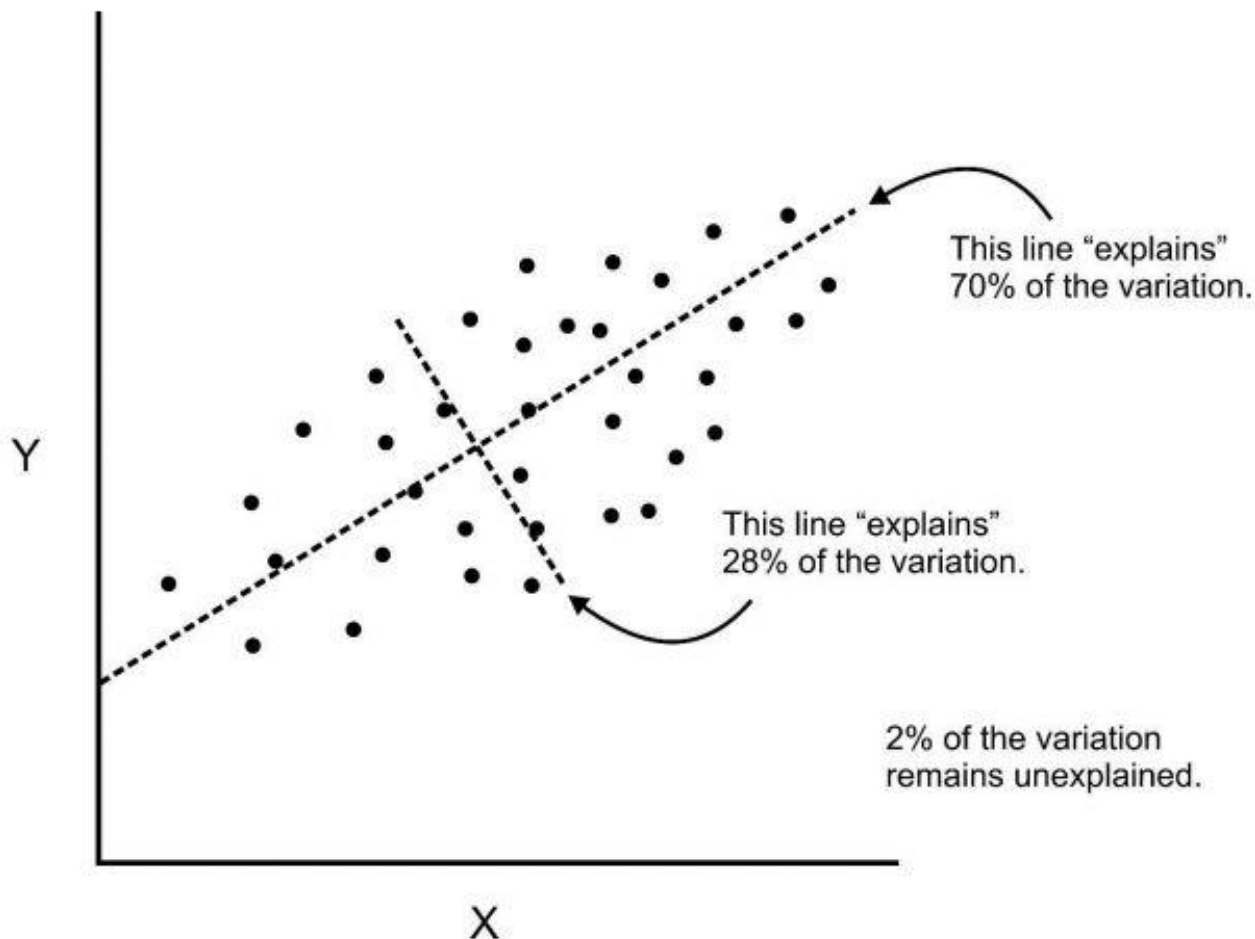
ما می توانیم یک خط متعامد اضافه کنیم.
اکنون می توانیم درک اجزا را شروع کنیم!



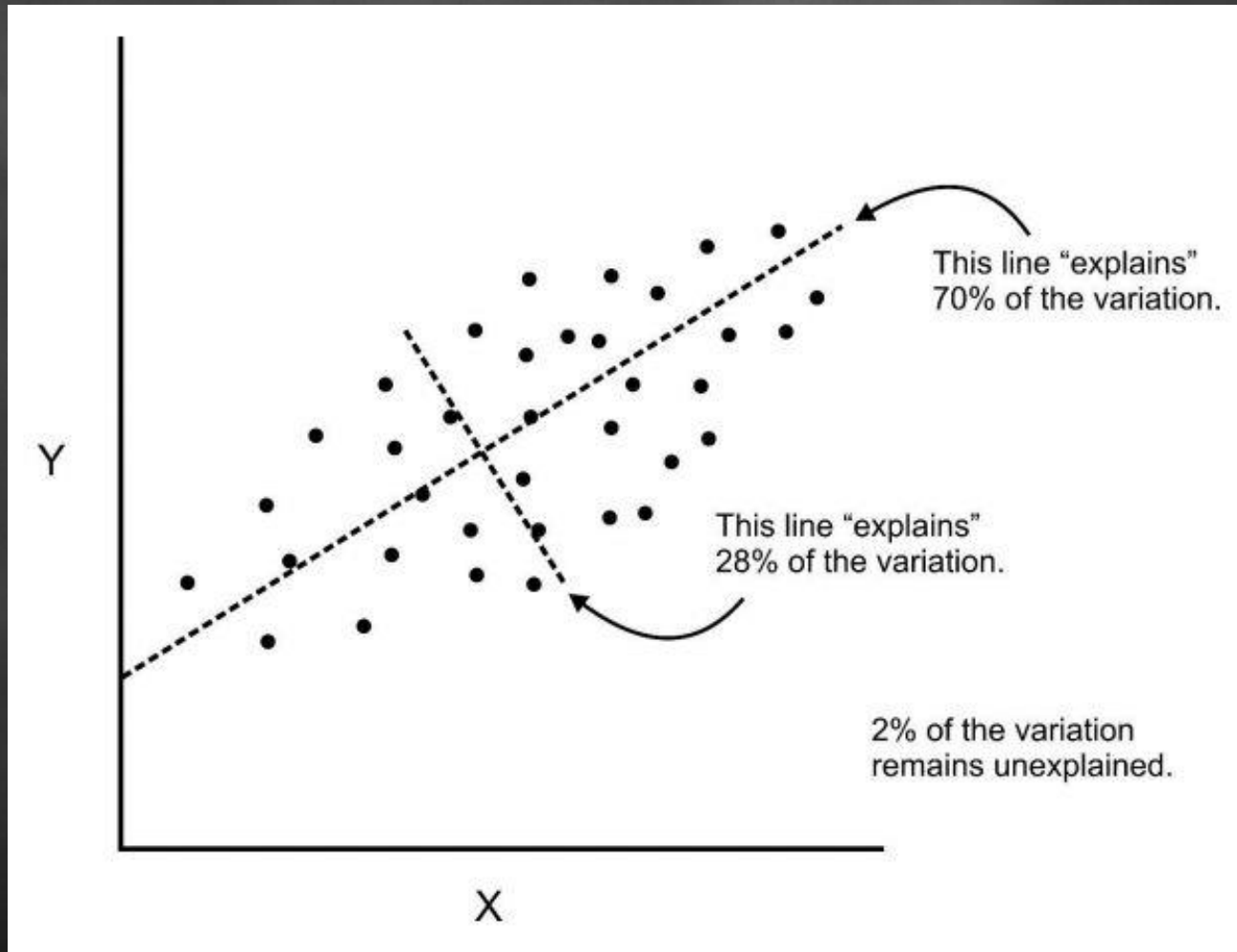
اجزاء یک رابطہ خطی دارند کہ یک سیستم متغیر را برای مجموعه داده انتخاب می کند به طوری کہ بیشترین واریانس مجموعه داده در محور اول قرار می گیرد.



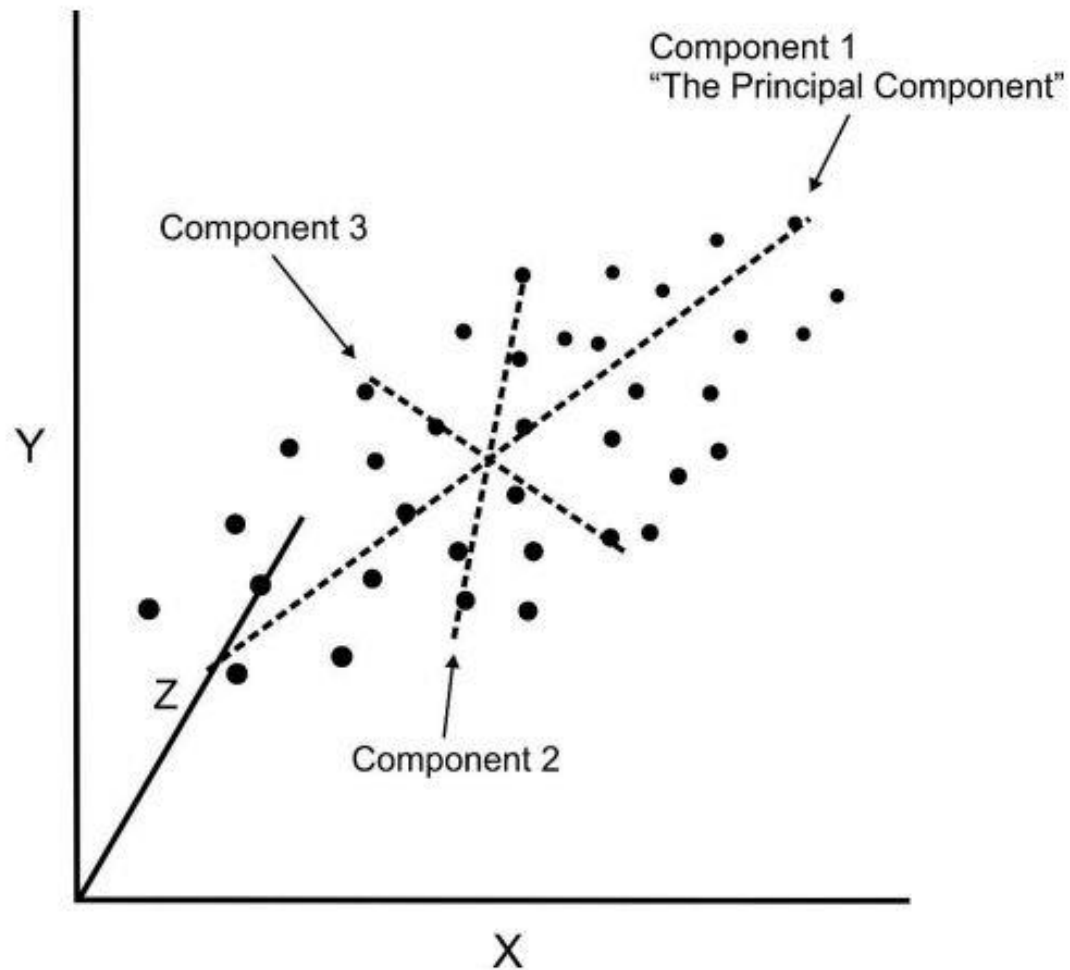
دومین واریانس بزرگ در محور دوم و غیره ... این فرایند به ما اجازه می دهد تا تعداد متغیرهای مورد استفاده در تحلیل را کاهش دهیم.



توجه داشته باشید که اجزاء همبستگی ندارند ، زیرا در فضای نمونه آنها با یکدیگر متعامد هستند.



ما می توانیم این تجزیه و تحلیل را در ابعاد بالاتر ادامه دهیم



اگر از این تکنیک در مجموعه داده با تعداد زیادی متغیر استفاده کنیم ، می توانیم مقدار تغییرات توضیح داده شده (ابعاد) را فقط به چند جزء فشرده کنیم. پ