Odbačene informacije

- Da smo sačuvali sve komponente $z_1, \ldots, z_D$, mogli bismo rekonstruisati originalni vektor x:

$$x = \sum_{i=1}^{D} z_i v_i$$

- Sa prvih K komponenti, rekonstrukcija je

$$\hat{x} = \sum_{i=1}^{K} z_i v_i$$

- Magnituda odbačenih informacija (greška rekonstrukcije) je:

$$\|x - \hat{x}\|^2 = \left\| \sum_{i=K+1}^{D} z_i v_i \right\|^2 = \sum_{i=K+1}^{D} z_i^2$$

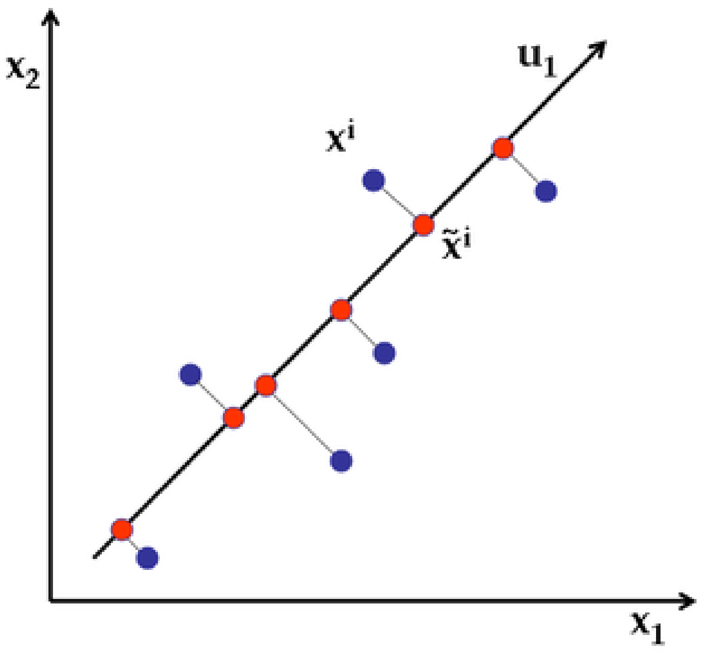
Odbačene informacije

- Novi koordinatni sistem je dobar ako je suma grešaka rekonstrukcije izračunata za sve primere skupa podataka mala ($\hat{x}_n \approx x_n$ za $n \in \{1, \dots, N\}$), tj., ako je malo:

$$\sum_{n=1}^{N} \|x_n - \hat{x}_n\|^2$$

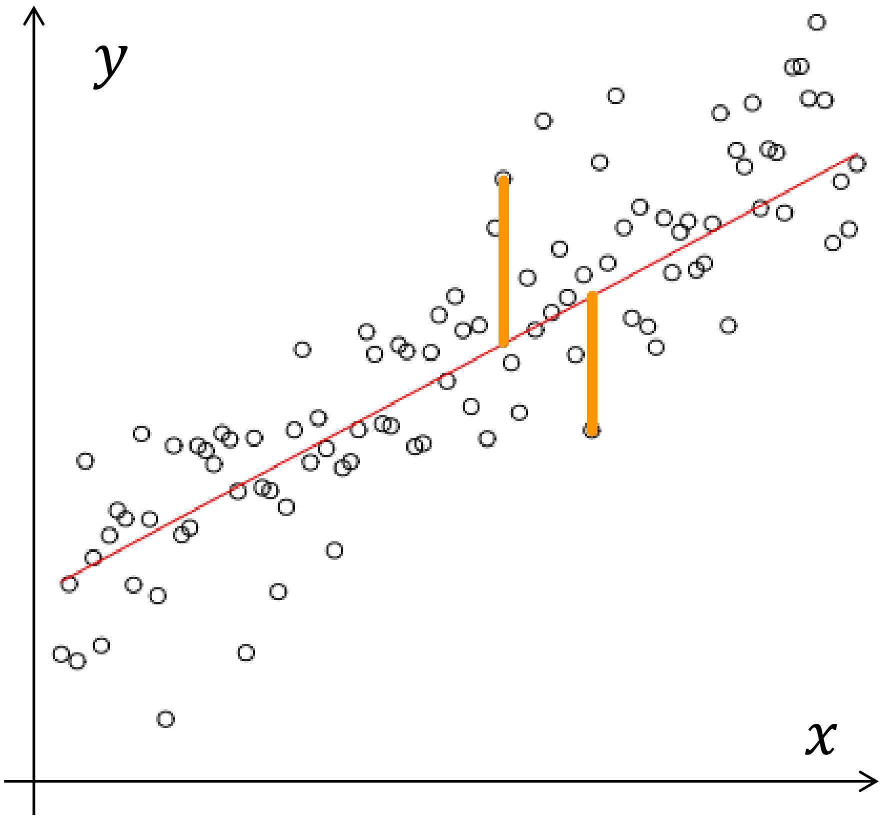
- PCA zato pronalazi koordinatni sistem koji mimimizuje ukupnu grešku rekonstrukcije

- Minimizujemo rastojanja tačaka od njihovih projekcija



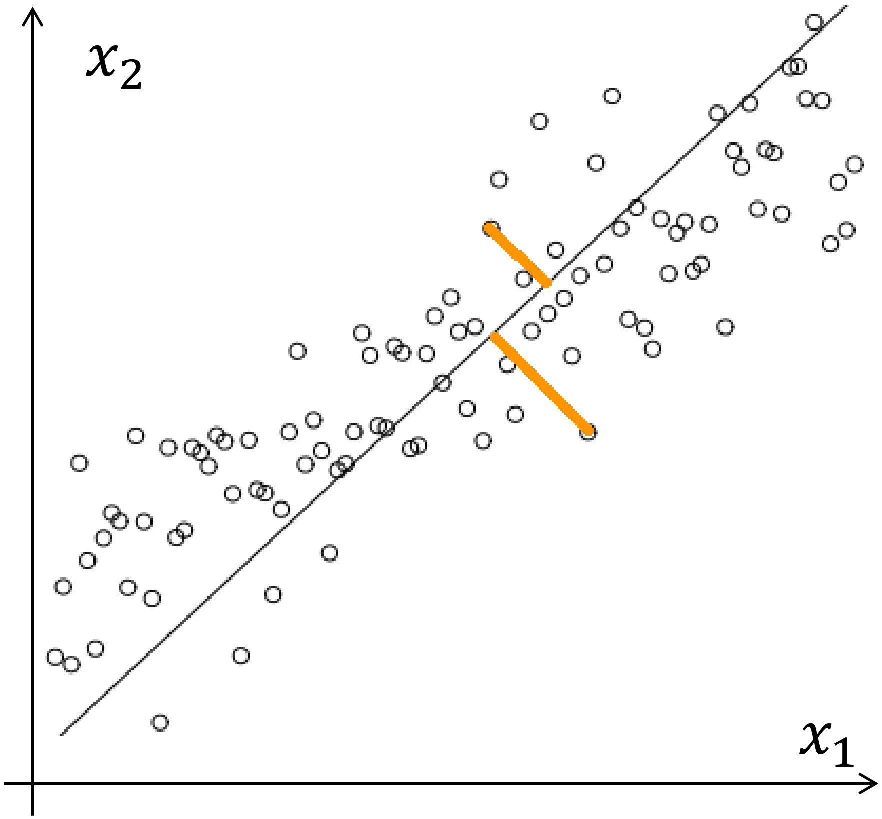
PCA *nije* linearna regresija

Linearna regresija



Specijalno obeležje y (ciljna varijabla) $y = f(x)$

PCA

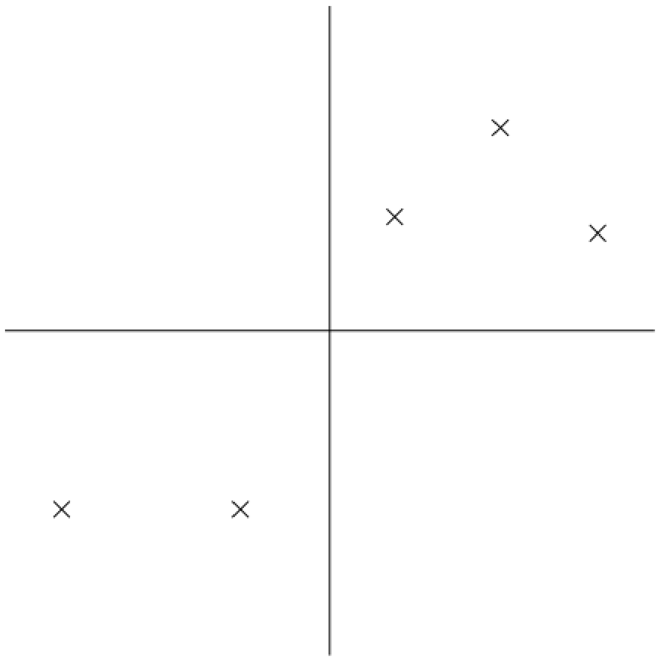


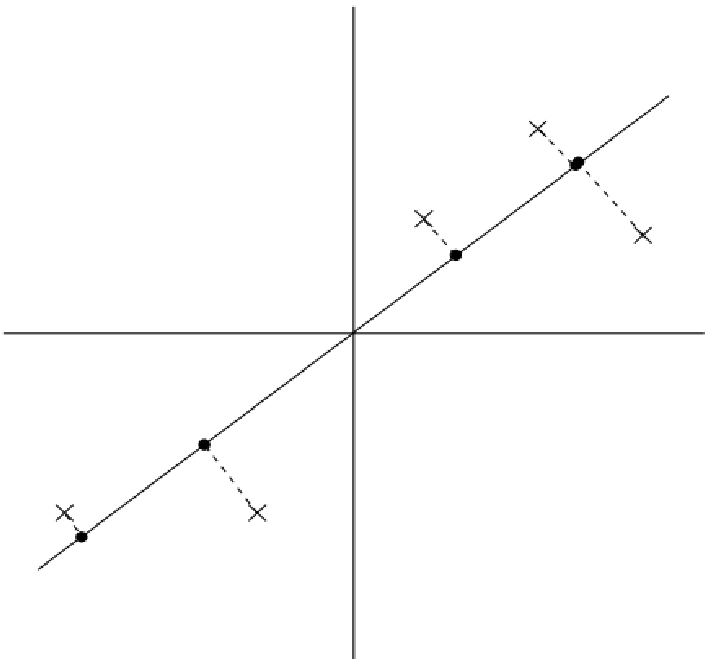
Sva obeležja su tretirana identično

PCA – druga (ekvivalentna) formulacija

- Cilj: zadržati što više informacija u podacima

  ⇒ novi koordinatni sistem određujemo tako da zadržimo što je moguće više *varijabilnosti* u dobijenoj projekciji



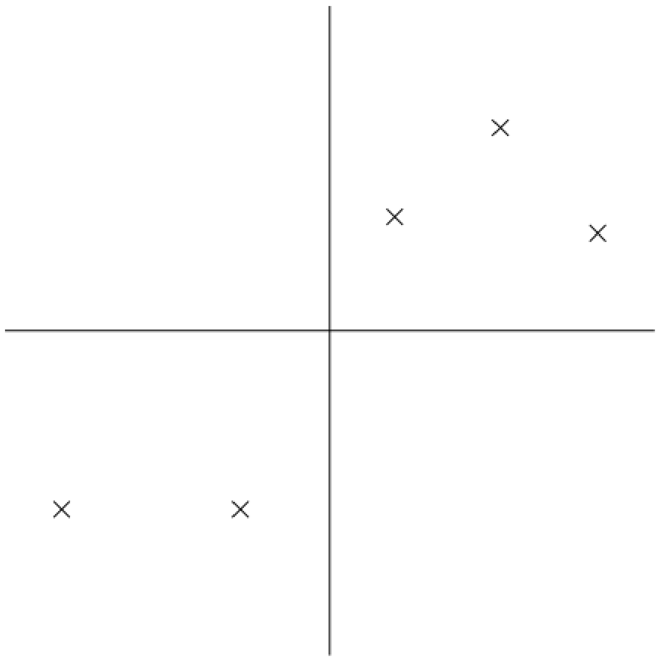


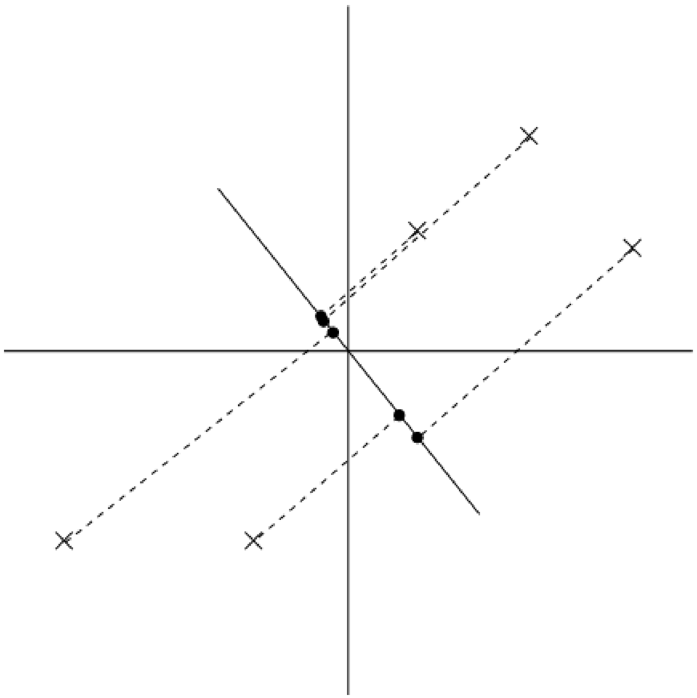
Projektovani podaci i dalje imaju dosta veliku varijansu. Podaci imaju tendenciju da budu daleko od nule

PCA – druga (ekvivalentna) formulacija

- Cilj: zadržati što više informacija u podacima

  $\Rightarrow$ novi koordinatni sistem određujemo tako da zadržimo što je moguće više *varijabilnosti* u dobijenoj projekciji

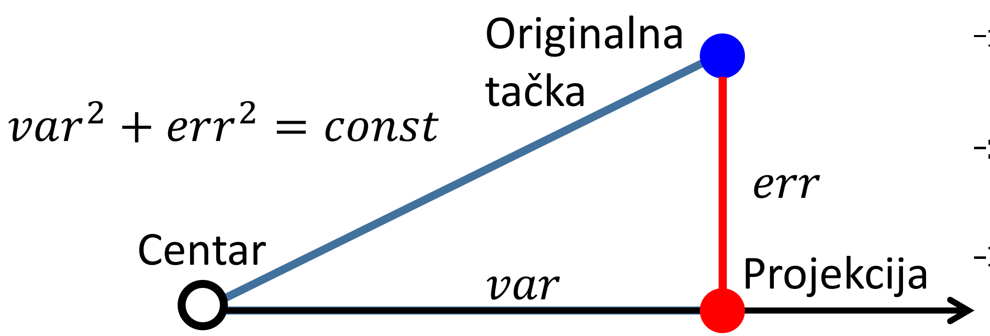


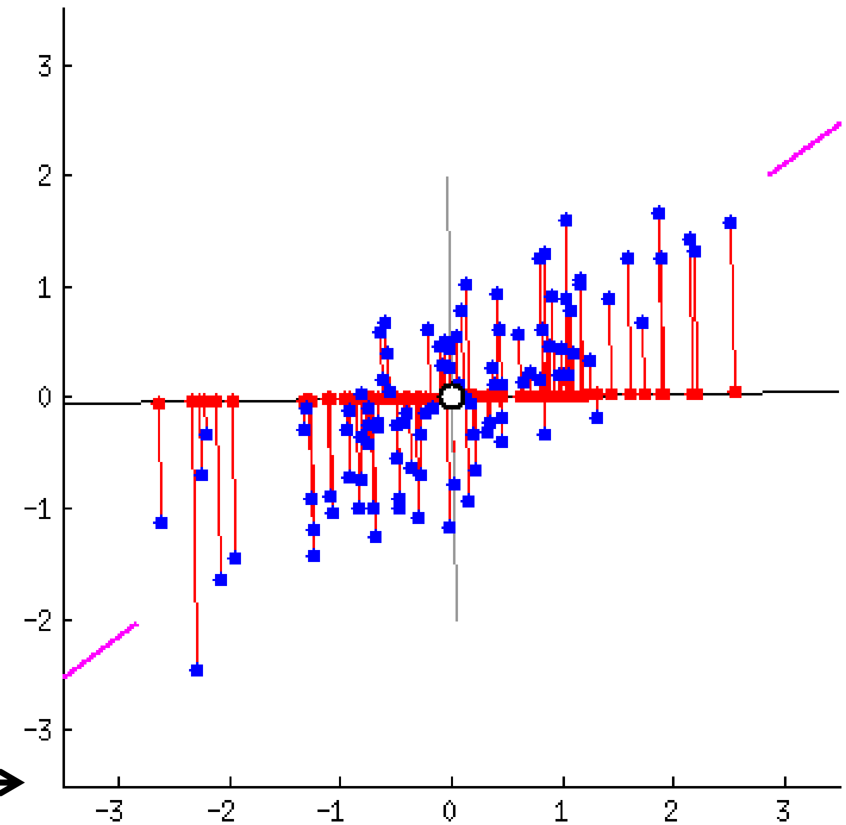


Projektovani podaci imaju znatno manju varijansu i mnogo su bliži nuli

PCA – dve ekvivalentne formulacije

- Ove dve formulacije PCA su ekvivalentne
  - Na slici možete primetiti da je zadržana varijansa projekcija („širina“ crvenih tačaka na novoj osi) najveća istovremeno kada je greška rekonstrukcije (suma crvenih linija) najmanja

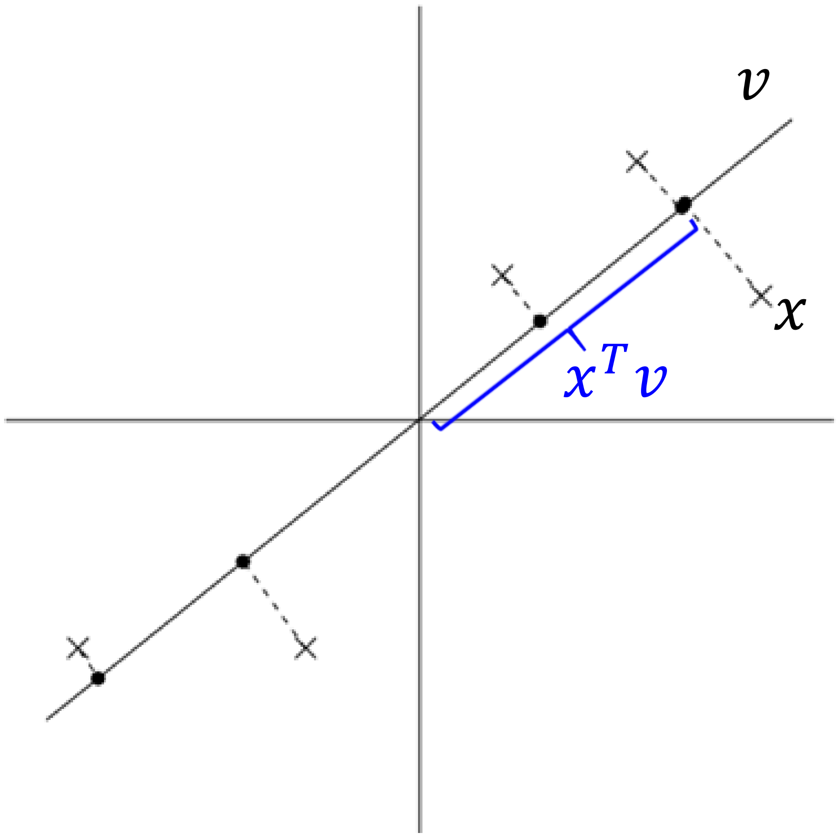




- PCA: Principal Component Analysis (analiza glavnih komponenti)
  - Nova osa je prva glavna komponenta

Principal Component Analysis (PCA)

- Želimo da automatski pronađemo pravac u koji će sačuvati najviše varijanse
- Dužina projekcije tačke x na *jedinični vektor* v je data sa $x^T v$:



- Dakle, da maksimizujemo varijansu projekcija, pronalazimo jedinični vektor v koji maksimizuje:

$$\frac{1}{N}\sum_{i=1}^{N}\left(\left(x^{(i)}\right)^{\mathrm{T}} v\right)^2$$

$$= \frac{1}{N}\sum_{i=1}^{N} v^{\mathrm{T}} x^{(i)}\left(x^{(i)}\right)^{\mathrm{T}} v$$

$$= v^{\mathrm{T}}\Bigg(\underbrace{\frac{1}{N}\sum_{i=1}^{N} x^{(i)}\left(x^{(i)}\right)^{\mathrm{T}}}_{\Sigma}\Bigg) v$$

Principal Component Analysis (PCA)

- Ako pretpostavimo da podaci imaju srednju vrednost 0, matrica

$$\Sigma = \frac{1}{N}\sum_{i=1}^{N} x^{(i)}\left(x^{(i)}\right)^{\mathrm{T}}$$

predstavlja empirijsku matricu kovarijanse podataka

Principal Component Analysis (PCA)

- Imamo optimizacioni problem:

$$\max_{v \in \mathbb{R}^D} v^{\mathrm{T}} \Sigma v \text{ pod ograničenjem } v^{\mathrm{T}} v = 1$$

za simetričnu matricu Σ

- Standardan način za rešavanje optimizacionih problema sa uslovima tipa jednakosti je formiranje *Lagranžijana*, ciljne funkcije koja u sebi sadrži ova ograničenja:

$$\mathcal{L}(x, \lambda) = \underbrace{v^T \Sigma v}_{\text{ciljna funkcija koju optimizujemo}} - \lambda \underbrace{v^T v}_{\text{ograničenje}}$$

λ – Lagranžov množilac („cena“ narušavanja ograničenja)

Principal Component Analysis (PCA)

- Da bi smo pronašli optimalno rešenje problema $\hat{v}$, izjednačićemo parcijalni izvod Lagranžijana po v sa 0:

$$\nabla_v \mathcal{L}(x, \lambda) = 2\Sigma^T v - 2\lambda v = 0$$
$$\Sigma v = \lambda v$$

- Sopstveni vektor (*eigenvector*) x i odgovarajuća sopstvena vrednost (*eigenvalue*) λ matrice A predstavljaju rešenja jednačine:

$$Ax = \lambda x$$

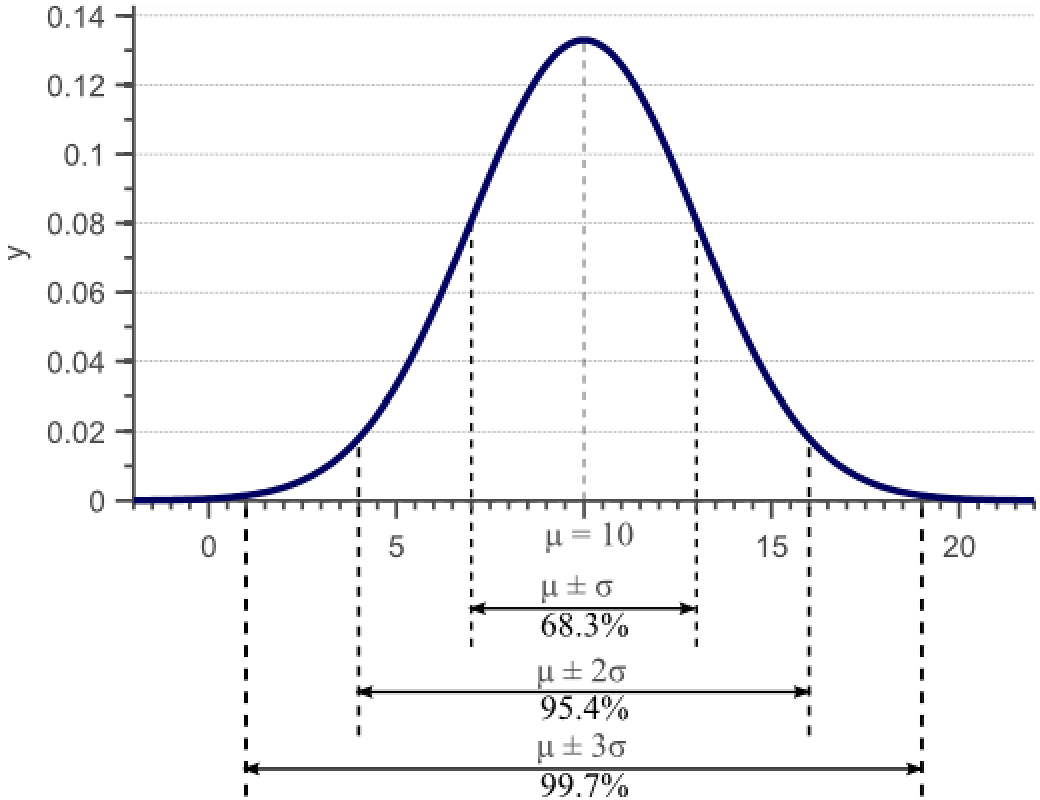
- Dakle, da bismo pronašli 1-dimenzioni potprostor sa kojim aproksimiramo podatke, treba da odaberemo v kao najveću glavnu komponentu (*eigenvector*) matrice Σ

Podsetnik – varijansa i kovarijansa

- Varijansa σ^2 meri „rasutost“ slučajne promenljive oko njene srednje vrednosti

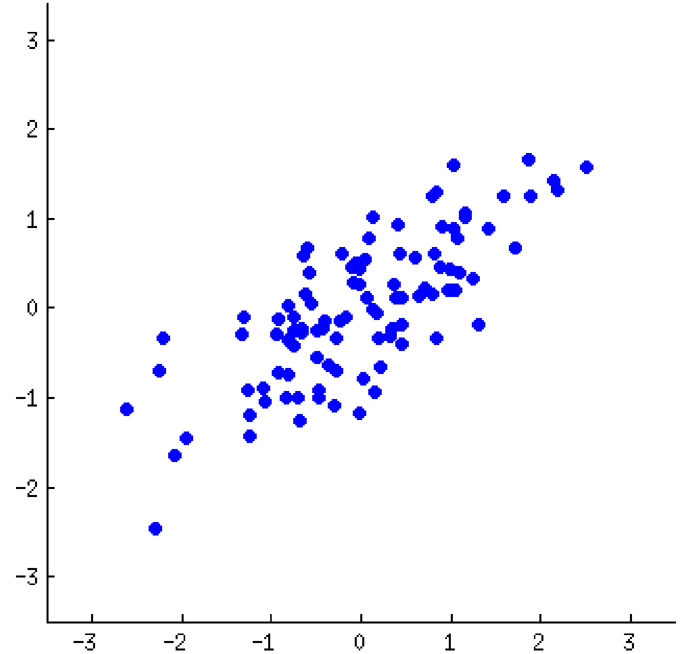
- ML ocena varijanse slučajne promenljive x iz uzorka:

$$\sigma_x^2 = \frac{1}{N}\sum_{i=1}^{N}\left(x^{(i)} - \mu\right)^2$$
$$= \sigma(x, x)$$



Podsetnik – varijansa i kovarijansa

- Međutim, varijansa se može koristiti samo da se opiše rasipanje podataka u pravcima paralelnim sa osama
- Ako pogledamo 2D primer na slici, možemo izračunati $\sigma(x, x)$ u pravcu x-ose i $\sigma(y, y)$ u pravcu y-ose. Ali, ovo horizontalno i vertiklano rasipanje nam ne objašnjavaju jasnu dijagonalnu korelaciju



$$\Sigma = \begin{bmatrix} 1.07 & 0.63 \\ 0.63 & 0.64 \end{bmatrix}$$

- Zato koncept varijanse proširujemo konceptom kovarijanse:

$$\sigma(x, y) = \mathbb{E}[(x - \mathbb{E}[x])(y - \mathbb{E}[y])]$$

- Za 2D podatke, matrica kovarijanse je:

$$\Sigma = \begin{bmatrix} \sigma(x, x) & \sigma(x, y) \\ \sigma(y, x) & \sigma(y, y) \end{bmatrix}$$

- Matrica Σ je simetrična (x je korelirano sa y isto koliko je y korelirano sa x)

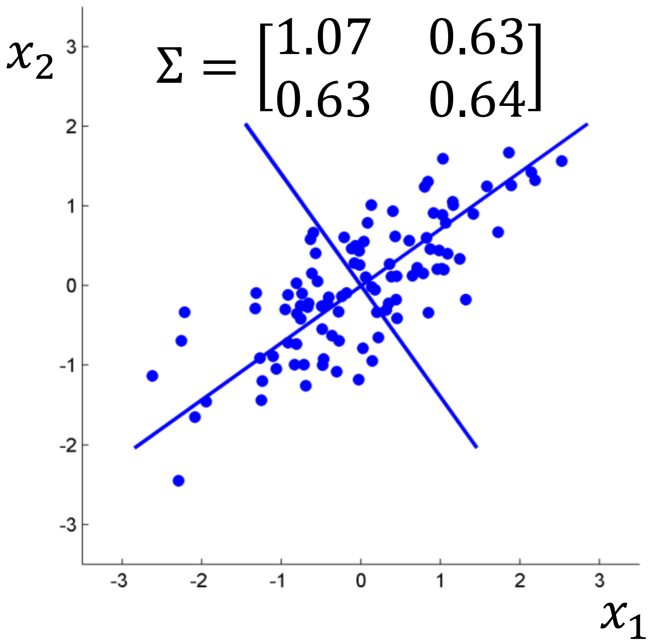
Sopstveni vektori matrice Σ

- Σ definiše i „rasipanje“ (varijansu) po osama i „orijentaciju“ (kovarijansu) podataka
- Sopstveni vektori (i njima odgovarajuće sopstvene vrednosti) jedinstveno definišu matricu Σ i time „oblikuju“ naše podatke
- Želimo da reprezentujemo Σ vektorom i magnitudom tako da:
  - Vektor pokazuje u pravcu najveće varijanse (najvećeg rasipanja) podataka
  - Magnituda ovog vektora pokazuje varijansu (rasipanje) u ovom pravcu

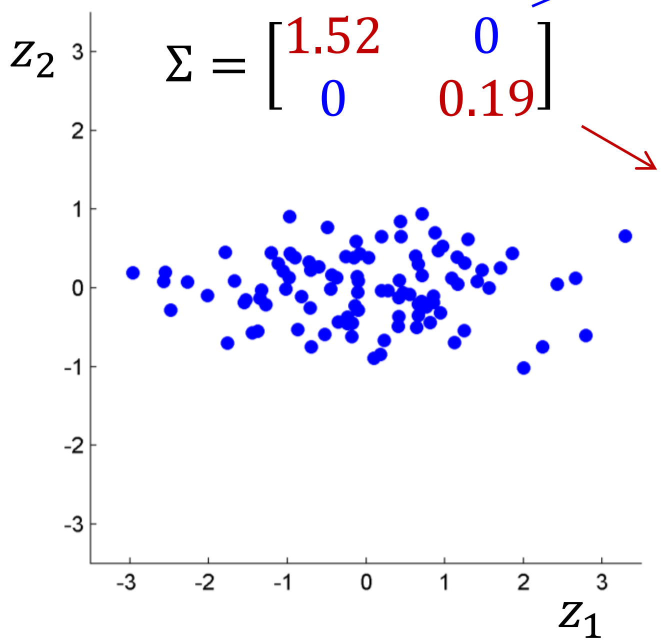
Sopstvene vrednosti matrice Σ

- Kvadratnu simetričnu matrica Σ transformišemo u dijagonalnu matricu
  - $\Sigma \in \mathbb{R}^{D \times D}$ ima D sopstvenih vektora $v_1, v_2, \dots, v_D$ kojima odgovaraju sopstvene vrednosti $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_D$
  - Sopstveni vektori biće ose novog koordinatnog sistema
  - U novom koordinatnom sistemu, odgovarajuće sopstvene vrednosti se nalaze na dijagonali transformisane matrice Σ

$$\Sigma = \begin{bmatrix} 1.07 & 0.63 \\ 0.63 & 0.64 \end{bmatrix}$$



$$\Sigma = \begin{bmatrix} 1.52 & 0 \\ 0 & 0.19 \end{bmatrix}$$



Korelacija z_1 i z_2 u novom koordinatnom sistemu je 0

Sopstvene vrednosti

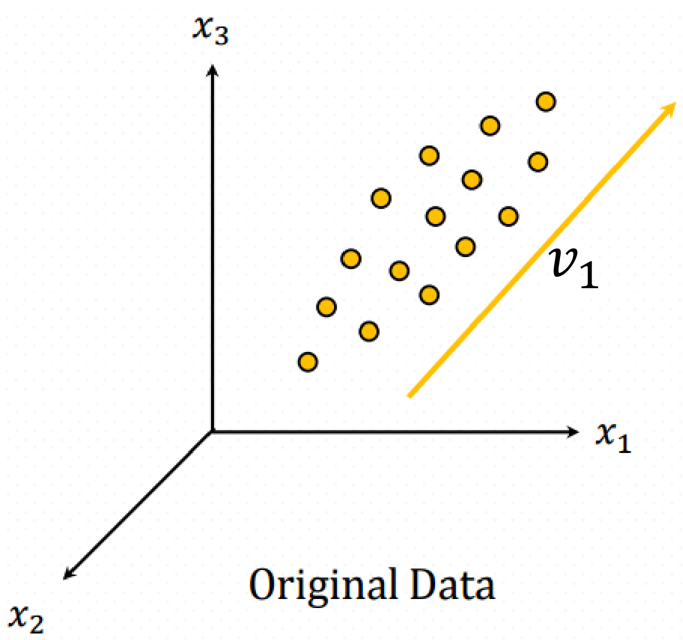
3Blue1Brown
2.83M subscribers

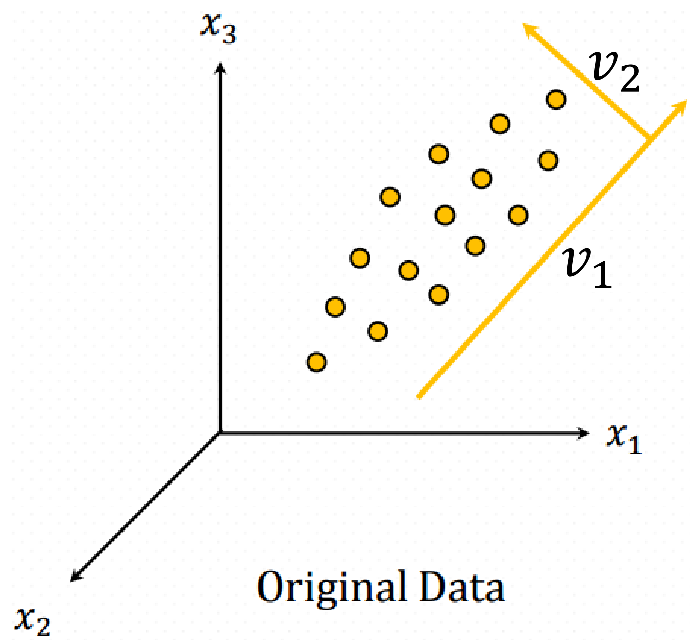
Sopstvene vrednosti matrice Σ

- **Teorema [Eckart-Young]**:
  - Pravci $v_1, v_2, \ldots, v_D$ rezultuju najboljom rekonstrukcijom podataka
  - Takođe, ovi vektori obuhvataju najviše varijanse

- Dakle, da bismo izračunali glavne komponente (ose novog koordinatnog sistema):
  - Pronađemo sopstvene vektore $v_1, v_2, \ldots, v_D$ matrice kovarijansi Σ kojima redom odgovaraju sopstvene vrednosti $\lambda_1 \geq \lambda_2 \geq \ \ldots \geq \lambda_D$
  - i-toj glavnoj komponenti odgovara i-ti sopstveni vektor v_i. Iznos varijanse koju opisuje ta komponenta je λ_i
  - Zbir $\lambda_1 + \lambda_2 + \cdots + \lambda_D$ je ukupna varijansa skupa podataka

PCA: dekompozicija matrice Σ





Prvi sopstveni vektor v_1 (sa najvećim λ_1) je u pravcu najveće varijabilnosti podataka

Naredna projekcija v_2 (drugo po veličini λ_2) je u pravcu sledeće najveće varijabilnosti, pri čemu je ortogonalna na sve prethodne projekcije

(da nije ortogonalna, obuhvatala bi varijabilnost već obuhvaćenu prethodnim projekcijama)