PCA

Saturday, September 23, 2023 11:30 AM

Data ⟹ Cov Matrix (Variance) ⟶ Eigen values & corresponding eigen vectors computed

Eigen val 1, val 2, val 3

Descending ordered eigenvalues: Val 2 > Val 3 > Val 1

Eigenvalues are descendingly arranged

$C = \begin{bmatrix} & & \\ & & \\ & & \end{bmatrix}$ Normalized Eigen vector matrix C

C * Data ⟶ Principal Comp matrix scores

Variances of PCA Score Variables are nothing but Eigen values of Variance – Covariance matrix (descending order)

$X_1 \; X_2 \ldots X_m$

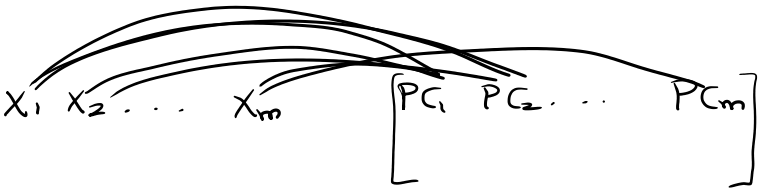
$PC_1 = X_1 a_1 + X_2 a_2 \ldots + X_m a_m$

$PC_2 = X_1 b_1 + X_2 b_2 + \ldots + X_m b_m$

$\vdots$

$PC_m = \_ \; \_ \; \_$

Note: All the Eigenvectors (Components) are orthogonal to each other.




```
print(prcomp.explained_variance_)
```

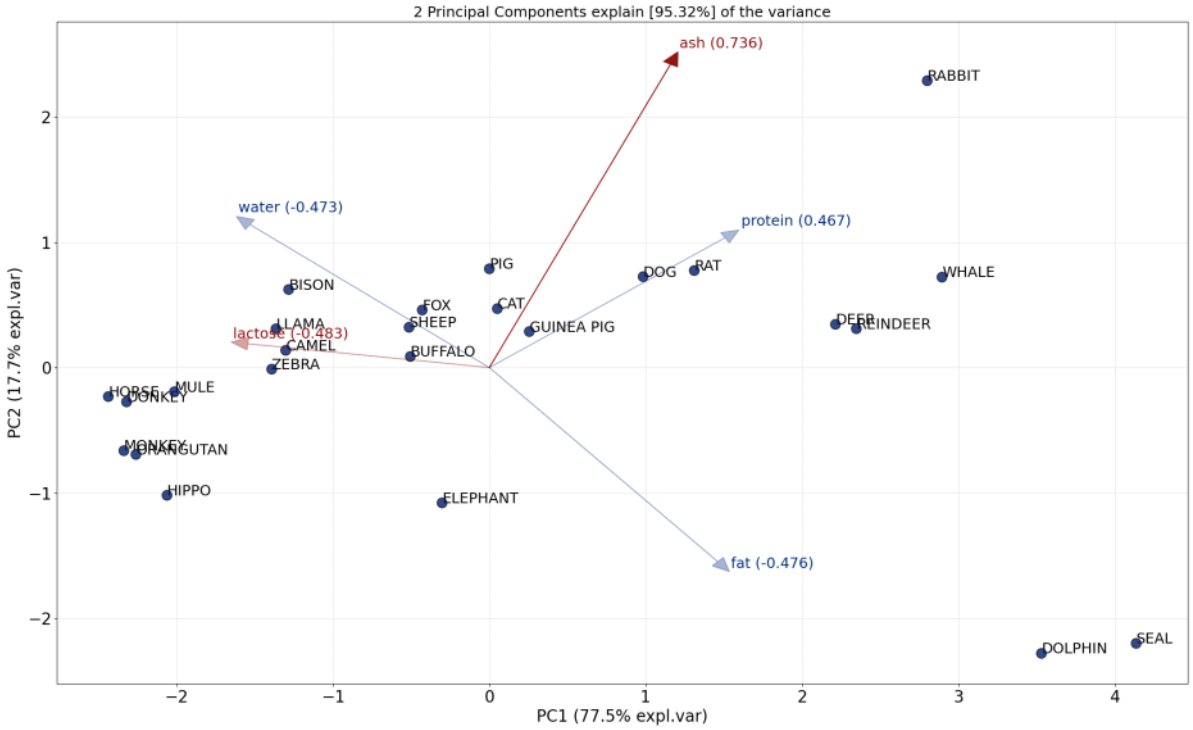
$Var(PC_1) > Var(PC_2) > \ldots > Var(PC_m)$

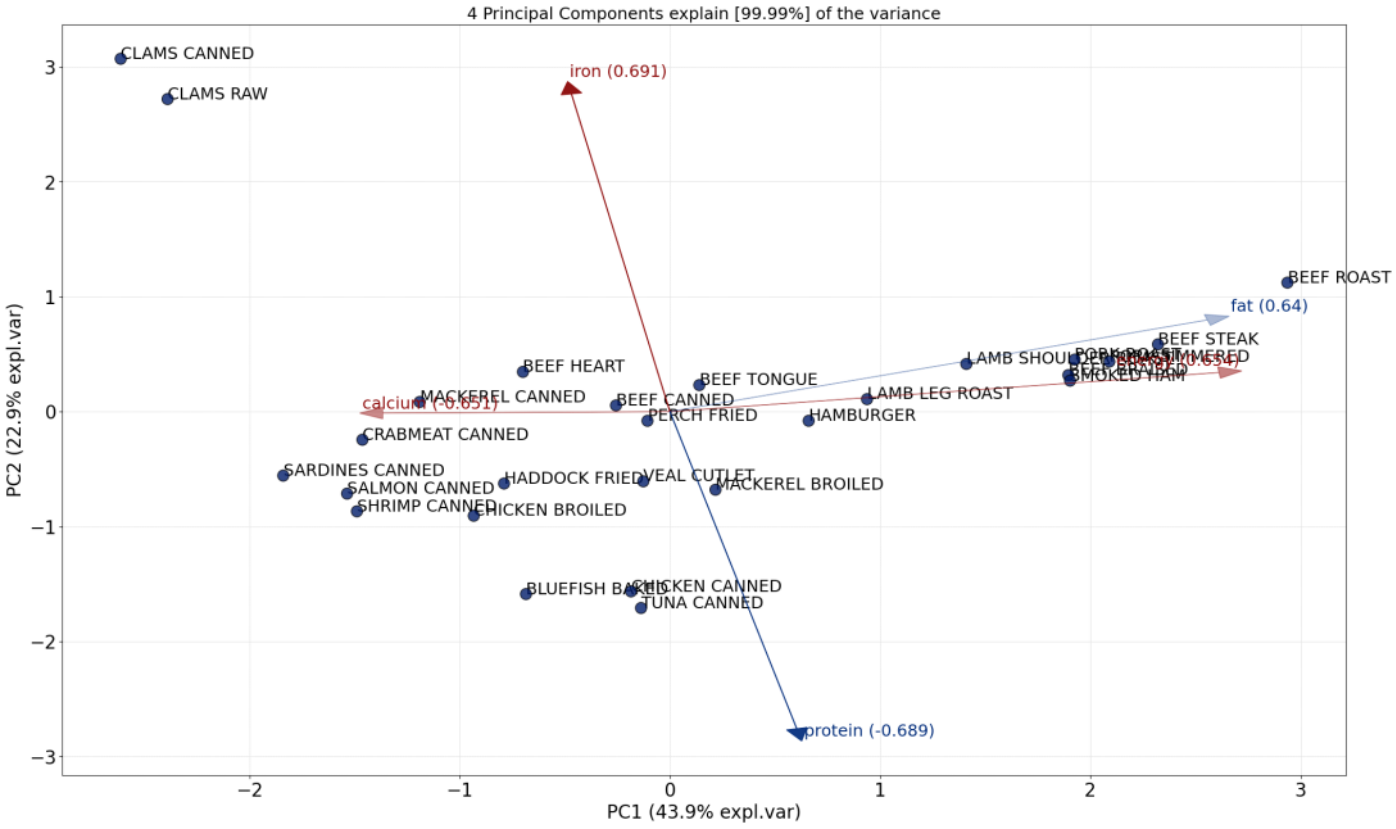
Eigen values of Variance – Covariance matrix

Good Practice: Scale the data, before you transform into PC scores.

$$\text{Total Variation} = Var(PC_1) + Var(PC_2) \ldots + Var(PC_m) = \sum_{i=1}^{m} Var(PC_i)$$

$$\text{Percentage of Variation explained by } j^{th} \text{ PC} = \frac{Var(PC_j)}{\sum_{i=1}^{m} Var(PC_i)} \times 100$$


2 Principal Components explain [95.32%] of the variance
ash (0.736)
water (-0.473)
protein (0.467)
lactose (-0.483)
fat (-0.476)
RABBIT
PIG
DOG
RAT
WHALE
BISON
FOX
CAT
LLAMA
SHEEP
GUINEA PIG
DEER
REINDEER
CAMEL
BUFFALO
ZEBRA
HORSE
MULE
DONKEY
MONKEY
ORANGUTAN
HIPPO
ELEPHANT
DOLPHIN
SEAL
PC2 (17.7% expl.var)
PC1 (77.5% expl.var)


4 Principal Components explain [99.99%] of the variance
CLAMS CANNED
CLAMS RAW
iron (0.691)
BEEF ROAST
fat (0.64)
BEEF STEAK
BEEF HEART
BEEF TONGUE
LAMB LEG ROAST
MACKEREL CANNED
BEEF CANNED
calcium (-0.651)
PERCH FRIED
HAMBURGER
CRABMEAT CANNED
SARDINES CANNED
HADDOCK FRIED
VEAL CUTLET
MACKEREL BROILED
SALMON CANNED
SHRIMP CANNED
CHICKEN BROILED
CHICKEN CANNED
TUNA CANNED
protein (-0.689)
PC2 (22.9% expl.var)
PC1 (43.9% expl.var)