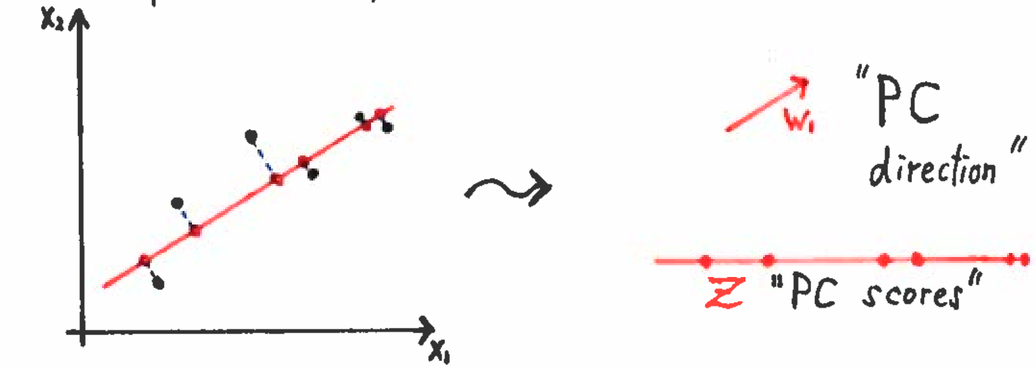
Principal Component Analysis



- From centred X via sample covariance matrix:

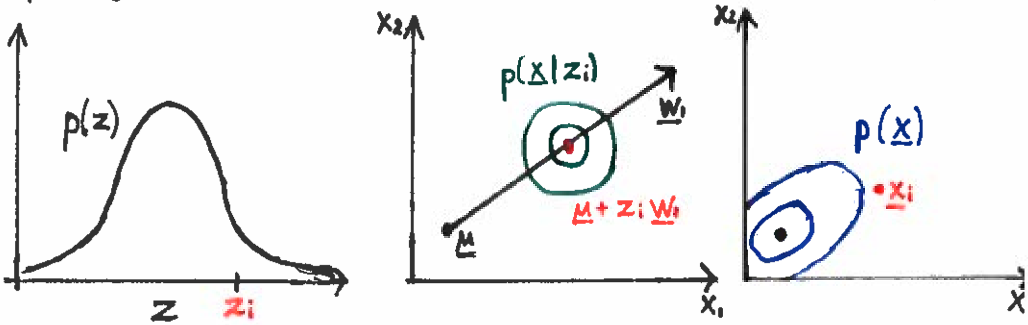
$$\Sigma = \frac{1}{n} X X^T = U \Lambda U^T; \quad Z = U_k^T X \quad ; \quad \underline{w}_i = \underline{u}_i$$

- From Gram matrix:

$$G = X^T X = V \tilde{\Lambda} V^T; \quad Z = \sqrt{\tilde{\Lambda}_k}\, V_k^T$$

Probabilistic PCA

Probabilistic model with latent variables



- PCA special case of Probabilistic PCA where covariance of noise is infinitesimally small

Quiz: Suppose we start from uncentred data.

a) Given uncentred data $\tilde{X}$, can we obtain corresponding PC scores Z?

b) Instead of the Gram matrix G, given inner products $\tilde{G} = \tilde{X}^T \tilde{X}$ between uncentred data, can we:

i) obtain corresponding PC scores Z?

ii) obtain underlying data $\tilde{\underline{x}}_i$? $\quad (\tilde{X}^T R^T)(\overset{\text{orth.}}{R} \cdot \tilde{X})$

c) Given all distances $\|\tilde{\underline{x}}_i - \tilde{\underline{x}}_j\|$, $i, j = 1, \dots, n$ between uncentred data points, can we

i) obtain corresponding Z?

ii) obtain underlying data $\hat{\underline{x}}_i$?

Dimensionality reduction

Preserve certain properties of data as much as possible in lower dim. space representation

PCA dimensionality reduction

- Represent the data as first k PC scores

1) Given uncentred observation matrix

$\tilde{X} = (\underline{\tilde{x}}_1 \dots \underline{\tilde{x}}_n)$

First centre: $X = \tilde{X} C_n$ where $C_n = I_n - \frac{1}{n} 1_n 1_n^T$

Then, decompose $\Sigma = \frac{1}{n} X X^T$ into EV decomp.

$\Sigma = U \Lambda U^T$ with sorted eigenvalues

PC scores: $Z = U_k^T X$ where $U_k = (\underline{u}_1, \dots, \underline{u}_k)$

$\rightsquigarrow$ X is $d \times n$ whereas Z is $k \times n$ (i.e. dim red.

If PC directions not needed, use Gram mat.

$G = X^T X$, EV-decomp. G into $V \tilde{\Lambda} V^T$

PC scores: $Z = \sqrt{\tilde{\Lambda}_k} V_k^T$

2) Given inner products between uncentred data $\tilde{G} = \tilde{X}^T \tilde{X}$

Compute Gram matrix G from $\tilde{G}$ and centring matrix C_n:

$$G = X^T X = C_n^T \tilde{X}^T \tilde{X} C_n = C_n \tilde{X}^T X C_n$$
$$= C_n \tilde{G} C_n$$

where $X = \tilde{X} C_n$ the centred data

"double centring": rows and columns have zero average

Calculate Z from G as before

- Matrices like G and $\tilde{G}$ called "similarity matrices"
- Possible to do dimensionality reduction given similarity matrices without having seen $\underline{x}_i$

3) Given squared distances between uncentred data points:

$$\delta_{ij}^2 = \|\underline{\tilde{x}}_i - \underline{\tilde{x}}_j\|^2 = (\underline{\tilde{x}}_i - \underline{\tilde{x}}_j)^T (\underline{\tilde{x}}_i - \underline{\tilde{x}}_j)$$

Define "distance matrix" Δ with elements

$$(\Delta)_{ij} = \delta_{ij}^2$$

Goal: calculate Gram matrix G from Δ:

We can ignore data centring

$$\delta_{ij}^2 = \|\underline{\tilde{x}}_i - \underline{\tilde{x}}_j\|^2 = \|(\underline{x}_i + \underline{\mu}) - (\underline{x}_j + \underline{\mu})\|^2$$

$$= \|\underline{x}_i - \underline{x}_j\|^2 = (\underline{x}_i - \underline{x}_j)^T (\underline{x}_i - \underline{x}_j)$$

$$\Rightarrow \delta_{ij}^2 = \underbrace{\|\underline{x}_i\|^2}_{\text{constant along row } i} + \|\underline{x}_j\|^2 - 2\underbrace{\underline{x}_i^T \underline{x}_j}_{\text{inner products}}$$

$\rightsquigarrow$ remove with C_n

We know: $(\Delta C_n)_{ij} = (\Delta)_{ij} - \frac{1}{n}\sum_{j=1}^{n} (\Delta)_{ij}$

"C_n from right removes row mean"

Let's calculate second term:

$$\frac{1}{n}\sum_{j=1}^{n} (\Delta)_{ij} = \|\underline{x}_i\|^2 + \frac{1}{n}\sum_{j=1}^{n} \|\underline{x}_j\|^2 - 2\underline{x}_i^T \underbrace{\left(\frac{1}{n}\sum_{j=1}^{n} \underline{x}_j\right)}_{=0}$$

$$= \|\underline{x}_i\|^2 + \frac{1}{n}\sum_{j=1}^{n} \|\underline{x}_j\|^2$$

since $\bar{\underline{x}} = \underline{0}$

2020 | 9 (5)

$$\Rightarrow (\Delta C_n)_{ij} = (\Delta)_{ij} - \left(\|\underline{x}_i\|^2 + \frac{1}{n}\sum_{j=1}^{n}\|\underline{x}_j\|^2\right)$$

$$= \|\underline{x}_i\|^2 + \|\underline{x}_j\|^2 - 2\underline{x}_i^T\underline{x}_j - \|\underline{x}_i\|^2 - \frac{1}{n}\sum_{j=1}^{n}\|\underline{x}_j\|^2$$

$$= \underbrace{\|\underline{x}_j\|^2}_{} - 2\underline{x}_i^T\underline{x}_j - \underbrace{\frac{1}{n}\sum_{j=1}^{n}\|\underline{x}_j\|^2}_{}$$

constant along column j of Δ

$\rightarrow$ remove with C_n

$$(C_n \Delta C_n)_{ij} = \ldots = -2\underline{x}_i^T\underline{x}_j$$

$$\Rightarrow C_n \Delta C_n = -2G$$

$$\Rightarrow G = -\frac{1}{2} C_n \Delta C_n$$

Then compute Z from G as before.