Data Analytics
EEE 4774 & 6777

Module 3

Dimensionality Reduction

Spring 2022

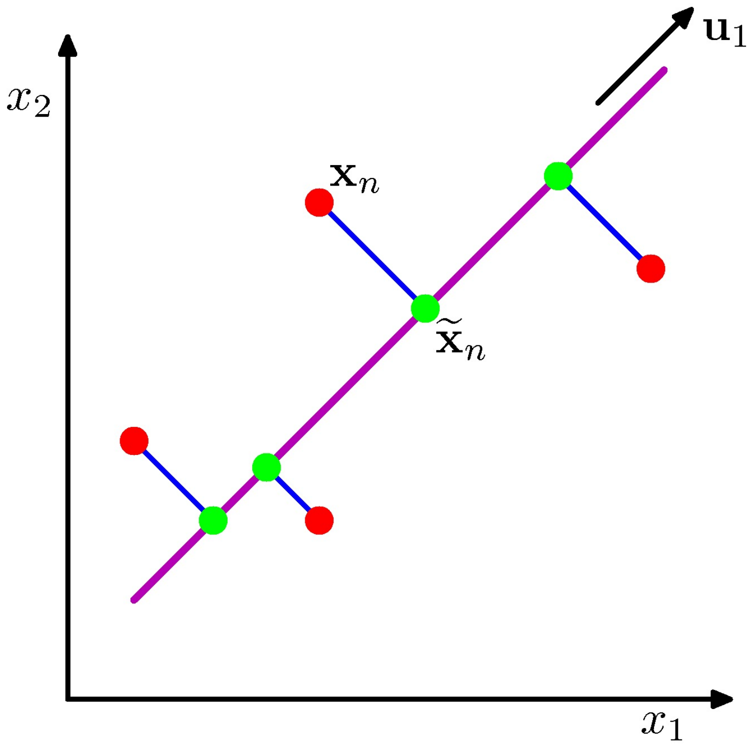
Principal Component Analysis (PCA)

- Used for
  - *dimensionality reduction,*
  - *feature extraction,*
  - *lossy data compression,*
  - *data visualization*

- Also known as Karhunen-Loeve Transform, Hotelling transform, proper orthogonal decomposition, etc. in different fields.

- **Definition 1:** orthogonal projection of data onto a lower dimensional linear space (principal subspace) such that the ***variance of the projected data is maximized***

- **Definition 2:** linear projection that ***minimizes the mean squared distance between the data points and their projections***



PCA: Maximum Variance Formulation

$\boldsymbol{u}_1^T \boldsymbol{x}_n$ = projected instance n
$\boldsymbol{u}_1^T \overline{\boldsymbol{x}}$ = projected mean

$$\max_{\boldsymbol{u}_1} \frac{1}{N} \sum_{n=1}^{N} (\boldsymbol{u}_1^T \boldsymbol{x}_n - \boldsymbol{u}_1^T \overline{\boldsymbol{x}})^2 = \max_{\boldsymbol{u}_1} \boldsymbol{u}_1^T \boldsymbol{S} \boldsymbol{u}_1$$

1 $$\overline{\boldsymbol{x}} = \frac{1}{N} \sum_{n=1}^{N} \boldsymbol{x}_n, \qquad \boldsymbol{S} = \frac{1}{N} \sum_{n=1}^{N} (\boldsymbol{x}_n - \overline{\boldsymbol{x}})(\boldsymbol{x}_n - \overline{\boldsymbol{x}})^T$$ 2

$$\max_{\boldsymbol{u}_1} \boldsymbol{u}_1^T \boldsymbol{S} \boldsymbol{u}_1 \quad \text{s.t.} \quad \boldsymbol{u}_1^T \boldsymbol{u}_1 = 1$$

$$\max_{\boldsymbol{u}_1} \boldsymbol{u}_1^T \boldsymbol{S} \boldsymbol{u}_1 + \lambda_1 (1 - \boldsymbol{u}_1^T \boldsymbol{u}_1)$$

$\frac{\partial}{\partial \boldsymbol{u}_1} = 0 \quad \Rightarrow \quad \boldsymbol{S} \boldsymbol{u}_1 = \lambda_1 \boldsymbol{u}_1 \quad \Rightarrow \quad$ $\boldsymbol{u}_1$ is an eigenvector of $\boldsymbol{S}$

$\boldsymbol{u}_1^T \boldsymbol{S} \boldsymbol{u}_1 = \lambda_1 \quad \Rightarrow \quad$ $\boldsymbol{u}_1$ is the eigenvector with the largest eigenvalue

3 $\boldsymbol{u}_1, \dots, \boldsymbol{u}_M$ eigenvectors corresponding to the M largest eigenvalues

PCA: Summary

- **Objective:** Learn new coordinates to reduce dimensions or extract features
- **Procedure:**
  - Compute sample mean and **covariance** $\boldsymbol{S}$
  - Compute first few eigenvectors of the sample covariance (new coordinates)

$$\boldsymbol{u}_1, \dots, \boldsymbol{u}_M \text{ where } M \ll D$$

  - Project data points onto the new coordinates to obtain lower dimensional representations

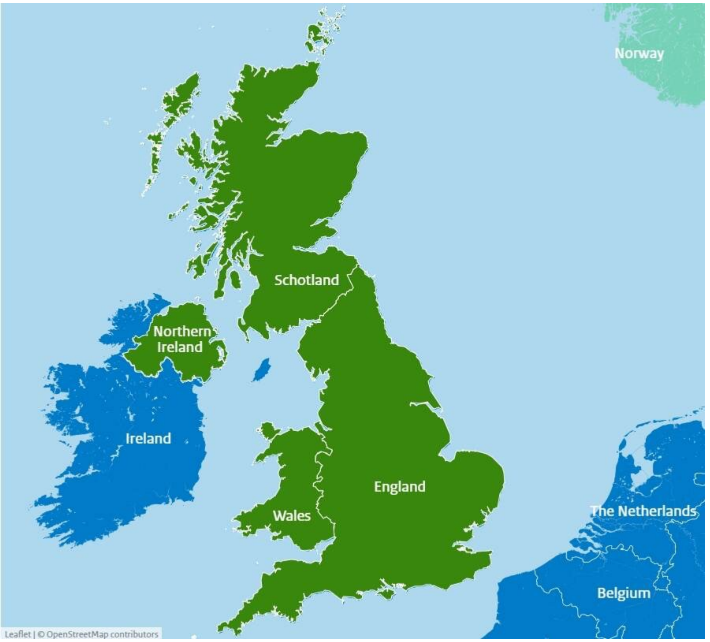
$$\tilde{x}_{n1} = \boldsymbol{u}_1^T \boldsymbol{x}_n$$

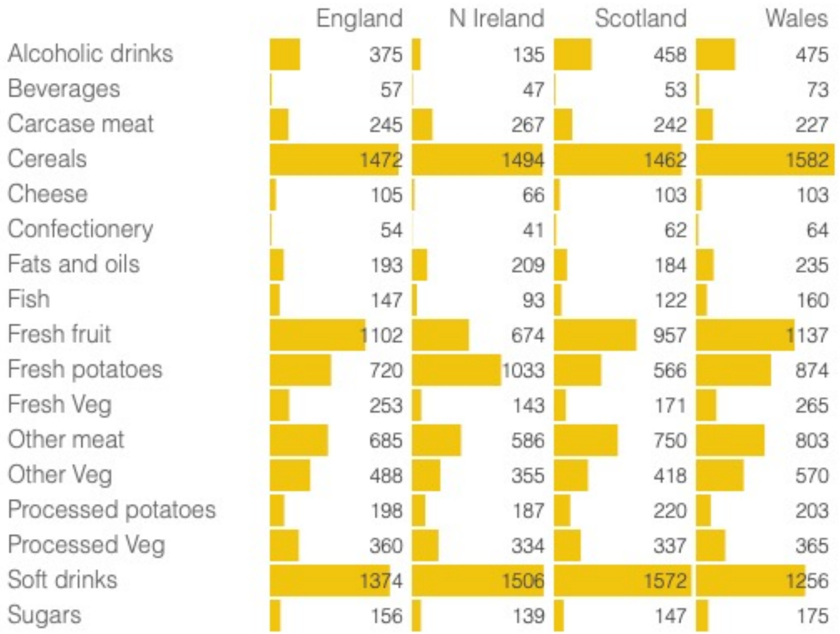
$$\tilde{\boldsymbol{x}}_n \in \mathbb{R}^M \text{ vs. } \boldsymbol{x}_n \in \mathbb{R}^D$$

Python Exercise: PCA Demo

- Data from UK Department for Environment, Food and Rural Affairs’ (DEFRA)
- Shows the consumption in grams (per person, per week) of 17 different types of foodstuff measured and averaged in the four countries of the United Kingdom in 1997
- 17 food types are the variables (features) and the 4 countries are the observations.

Country	Alcoholic drinks	Beverages	Carcase meat	Cereals	Cheese	Confectionery	Fats and oils	Fish	Fresh fruit	Fresh potatoes	Fresh Veg	Other meat	Other Veg	Processed potatoes	Processed Veg	Soft drinks	Sugars
England	375	57	245	1472	105	54	193	147	1102	720	253	685	488	198	360	1374	156
N Ireland	135	47	267	1494	66	41	209	93	674	1033	143	586	355	187	334	1506	139
Scotland	458	53	242	1462	103	62	184	122	957	566	171	750	418	220	337	1572	147
Wales	475	73	227	1582	103	64	235	160	1137	874	265	803	570	203	365	1256	175

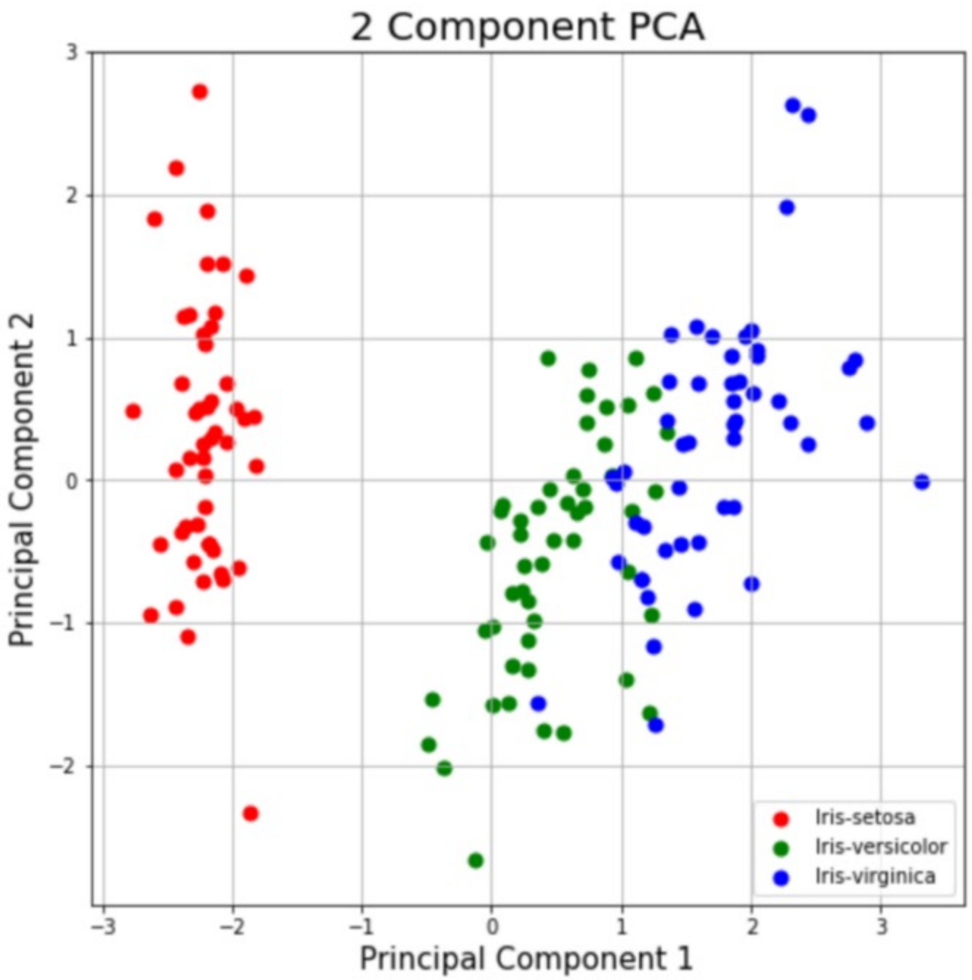




PCA Demo

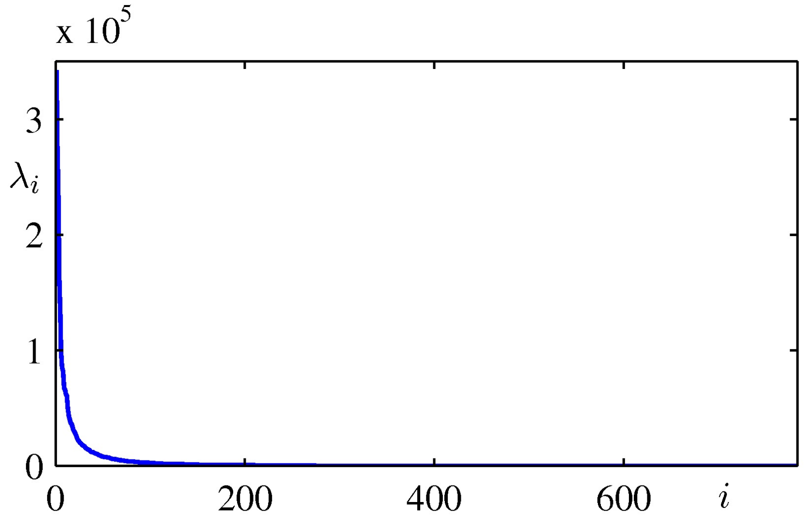
- Visualization of Iris Data:
https://towardsdatascience.com/pca-using-python-scikit-learn-e653f8989e60

- UK food data
http://setosa.io/ev/principal-component-analysis/



Selecting Number of Dimensions

- Eigenvalue scree plot
(may not be seen in practice)



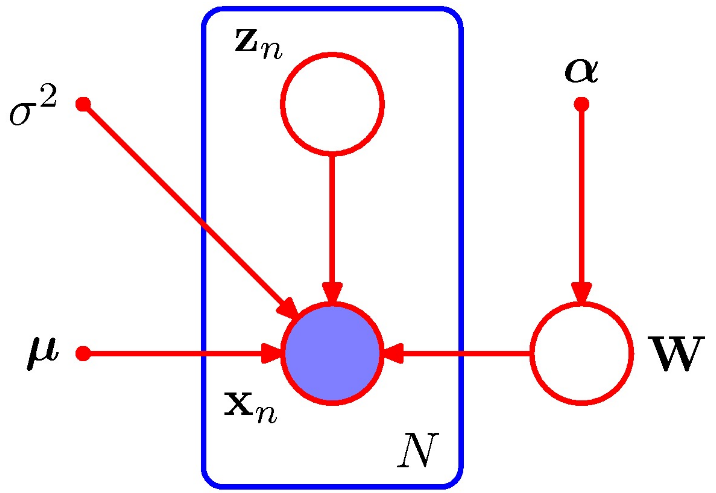
(a)

- Cross-validation
(computationally costly)

- Bayesian PCA
(approximations needed)

$$p(\boldsymbol{W}|\boldsymbol{\alpha}) = \prod_{i=1}^{M} \left(\frac{\alpha_i}{2\pi}\right)^{D/2} e^{-\alpha_i \boldsymbol{w}_i^T \boldsymbol{w}_i/2}$$

$$p(\boldsymbol{X}|\boldsymbol{\alpha}, \boldsymbol{\mu}, \sigma^2) = \int p(\boldsymbol{X}|\boldsymbol{W}, \boldsymbol{\alpha}, \boldsymbol{\mu}, \sigma^2)\, p(\boldsymbol{W}|\boldsymbol{\alpha})\, \mathrm{d}\boldsymbol{W}$$



Probabilistic PCA

$$\boldsymbol{x} = \boldsymbol{W}\boldsymbol{z} + \boldsymbol{\mu} + \boldsymbol{\epsilon}$$

$$p(\boldsymbol{x}|\boldsymbol{z}) = \mathcal{N}(\boldsymbol{x}|\boldsymbol{W}\boldsymbol{z} + \boldsymbol{\mu}, \sigma^2 \boldsymbol{I})$$

$$p(\boldsymbol{z}) = \mathcal{N}(\boldsymbol{z}|\boldsymbol{0}, \boldsymbol{I})$$

$$p(\boldsymbol{x}) = \int p(\boldsymbol{x}|\boldsymbol{z})\, p(\boldsymbol{z})\, d\boldsymbol{z}$$

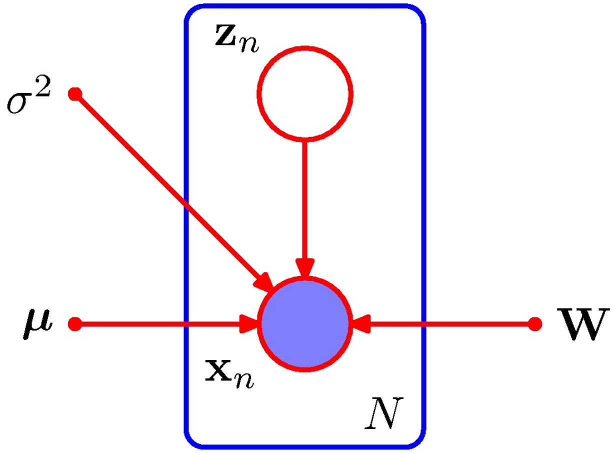
$$= \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{C})$$

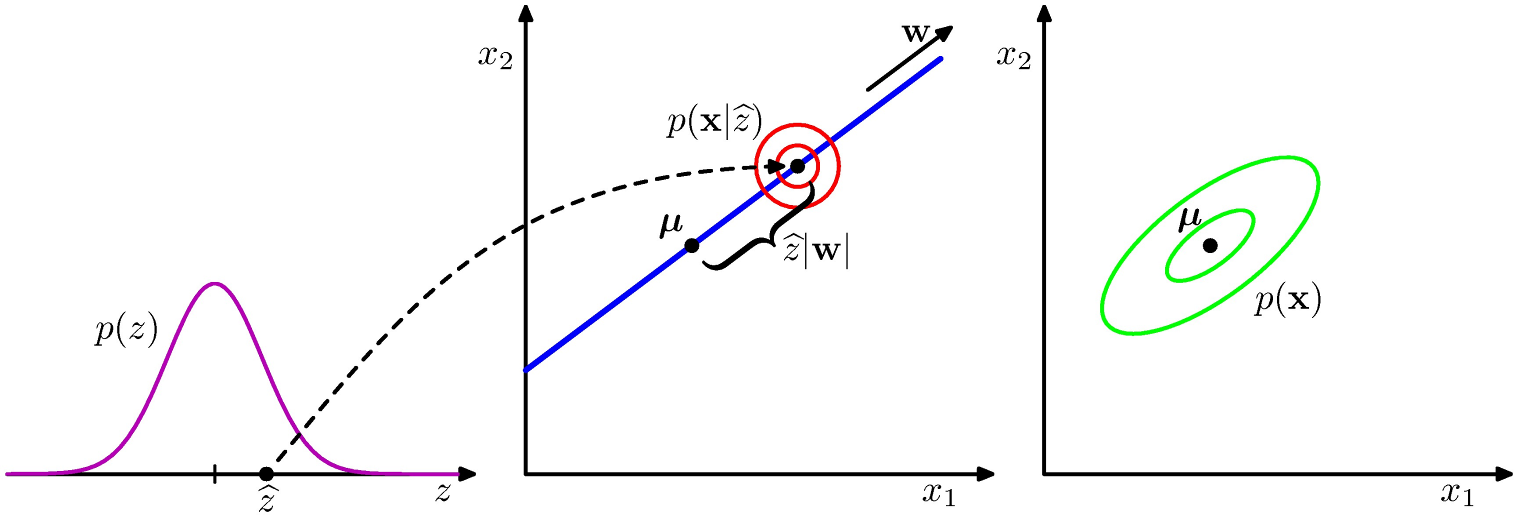
$$\boldsymbol{C} = \boldsymbol{W}\boldsymbol{W}^T + \sigma^2 \boldsymbol{I}$$

$$\max_{\boldsymbol{\mu}, \boldsymbol{W}, \sigma^2} \sum_{n=1}^{N} \log p(\boldsymbol{x}_n|\boldsymbol{\mu}, \boldsymbol{W}, \sigma^2)$$

$$\boldsymbol{\mu}_{ML} = \overline{\boldsymbol{x}} \qquad \sigma^2_{ML} = \frac{1}{D-M} \sum_{i=M+1}^{D} \lambda_i$$

$$\boldsymbol{W}_{ML} = \underbrace{\boldsymbol{U}_M}_{\text{Corresponding eigenvectors}} (\underbrace{\boldsymbol{L}_M}_{\text{Diagonal with M largest eigenvalues of } \boldsymbol{S}} - \sigma^2 \boldsymbol{I})^{1/2} \underbrace{\boldsymbol{R}}_{\text{Arbitrary orthogonal}}$$





Factor Analysis

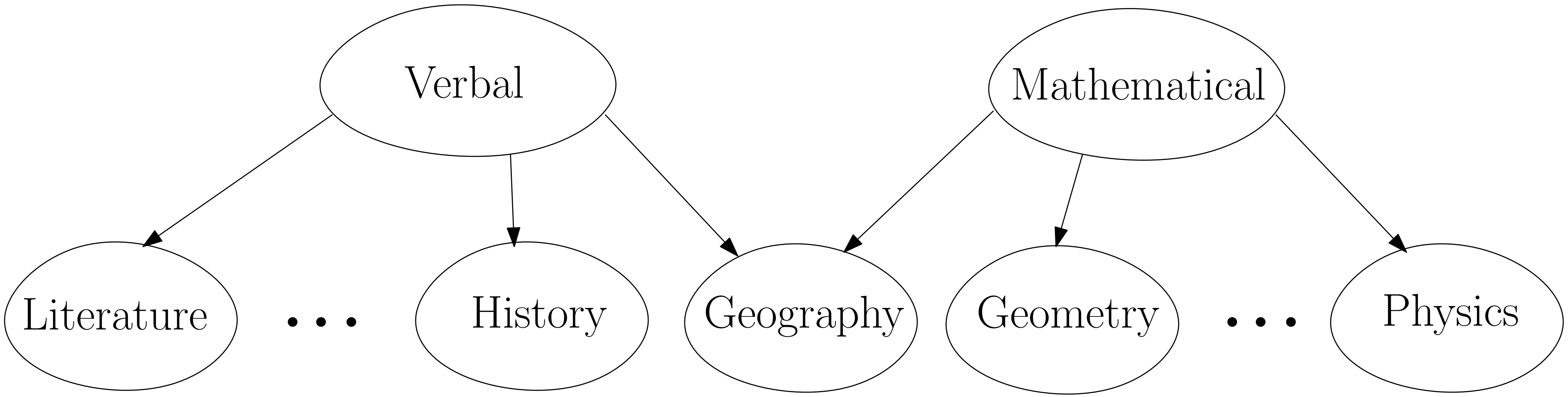
$$p(\boldsymbol{x}|\boldsymbol{z}) = \mathcal{N}(\boldsymbol{x}|\boldsymbol{W}\boldsymbol{z} + \boldsymbol{\mu}, \underbrace{\boldsymbol{\Psi}}_{\text{Diagonal}})$$ **(FA)** vs. **(PCA)** $$p(\boldsymbol{x}|\boldsymbol{z}) = \mathcal{N}(\boldsymbol{x}|\boldsymbol{W}\boldsymbol{z} + \boldsymbol{\mu}, \sigma^2\boldsymbol{I})$$

$$p(\boldsymbol{z}) = \mathcal{N}(\boldsymbol{z}|\boldsymbol{0}, \boldsymbol{I})$$

- Maximum likelihood: $\boldsymbol{\mu}_{ML} = \overline{\boldsymbol{x}}$
- No closed-form solution for $\boldsymbol{W}$, unlike probabilistic PCA
- Find iteratively using Expectation Maximization (EM) algorithm
- Similarly finds a new representation in a lower dimensional space
- Psychologist Charles Spearman (1904) used to explain the positive correlation of student success

in seemingly different subjects in terms of General intelligence factor (g factor)

Factor Analysis



Independent Component Analysis (ICA)

- Differently from PCA, the latent variable $\boldsymbol{z}$ is **non-Gaussian** and its entries are **independent**

- Used for *Blind Source Separation*

  - e.g., *Cocktail Party Problem:* a group of people talking at the same time, and multiple microphones picking up mixtures of independent sources.

  - Separate original sources when only mixed signals are available (blind)

  - Non-Gaussian and independent original sources are suitable assumptions for Cocktail Party Problem