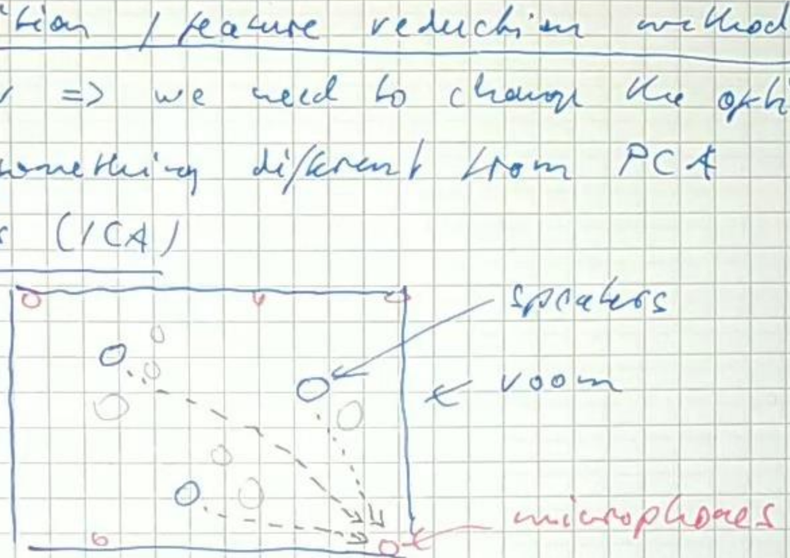
- alternative <u>linear decomposition / feature reduction methods</u>

$z = \varphi(x) = W \cdot x$ with φ linear $\Rightarrow$ we need to change the optimality criterion to get something different from PCA

<u>Independent Component Analysis (ICA)</u>

- cocktail party problem:



z_j: time series of voice of speaker j

X_i: time series of recording of microphone i

model: $X = W \cdot Z + \varepsilon$ ← noise

~~$i, j = 1, \ldots$~~ $i = 1 \ldots N$ # microph

$j = 1 \ldots D$ # speakers

$N \geq D$

$X_{it}, Z_{jt}, \; t = 1 \ldots T$ # time steps

W_{ij}: how much speaker j influences microphone i

w_{ij} big $\hat{=}$ j is near to i or speaks very loudly

w_{ij} small $\hat{=}$ j cannot be heared (almost) by microphone i

task: • recover, what everyone said (Z) from the mixed microphone recording (X) $\hat{=}$ "unmixing problem"

- easy, when matrix W is known: $Z = W^+ X$ $\quad W^+ = (W^T W)^{-1} W^T$ pseudo-inverse
- <u>but</u>: W is unknown, because we do not know the location of the speakers

$\Rightarrow$ need to recover Z and W^+ simultaneously

- idea: what each speaker says is statistically independent of every other speaker
  $\Rightarrow W^+$ must be orthonormal
  $\Rightarrow$ objective $\hat{W}^+, \hat{Z} = \underbrace{\arg\min_{W^+, Z} \sum_j}_{\text{maximizes independence between } Z_j \text{ and } Z_{j'}} H(Z_j)$ s.t. $Z = W^+ X$, W^+ orthonormal
  $H(Z_j)$ ← entropy of the time series Z_j
- mathematical proof: solvable, when ε (noise) is <u>not</u> Gaussian
  otherwise ambiguous $\hat{=}$ "non-identifiable"
- algorithm: Fast ICA [Hyvärinen & Oja, 2000]
  with lots of tricks to make fast and robust in practice
  (non-Gaussian noise here means: rewrite the ~~decomposition~~ model
  $X = WZ + \varepsilon = W(Z^* + \varepsilon')$
  time series $Z_j^* + \varepsilon_j$, if marginalized over t, has not a Gaussian distribution)

Non-negative Matrix Factorization (NMF)

- $X \geq 0$ ($\hat{=} X_{ij} \geq 0$) $\Rightarrow X = Z \cdot H$ with $Z \geq 0$, $H \geq 0$
- interpretation: customer preferences: big $X_{ij} \hat{=}$ customer i likes item j
  rows of H are "topics" or "categories" big $H_{kj} \hat{=}$ item j belongs to category k
  $k = 1, \dots, C$,
  example: items are movies $k \in \{$comedy, drama, action, ...$\}$
  rows of Z are "loadings": big $Z_{ik} \hat{=}$ customer i likes topic k

$\Rightarrow$ recommender system: if we find out z_{ik} is big, recommend items from category k to customer i in the hope that he/she will buy

- solution by alternating optimization:

  ⓪ find initial guesses $Z^{(0)}$ and $H^{(0)}$ (with non-negative entries)

  ① repeat $t = 1, \dots, T$ or until convergence

  ⓐ optimize H with Z fixed:

$$H^{(t)} = \arg\min_H \frac{1}{2} \| X - Z^{(t-1)} H \|^2$$

  ⓑ optimize Z with H fixed

$$Z^{(t)} = \arg\min_Z \frac{1}{2} \| X - Z \cdot H^{(t)} \|^2$$

- algorithm: optimize using gradient descent, where the learning rate is a cleverly chosen function of the current guesses $Z^{(t-1)}, H^{(t-1)}$

$\Rightarrow$ multiplicative updates

$$H^{(t)} = H^{(t-1)} * \frac{(Z^{(t-1)})^T X}{(Z^{(t-1)})^T Z^{(t-1)} H^{(t-1)}}$$

$$Z^{(t)} = Z^{(t-1)} * \frac{X (H^{(t)})^T}{Z^{(t-1)} H^{(t)} (H^{(t)})^T}$$

$*$ multiplication and — division elementwise

denominator and numerator are matrix multiplication