Probabilistic Archetypal Analysis

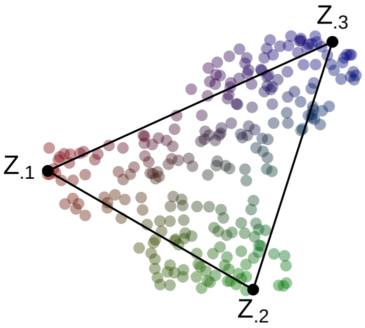
Sohan Seth

University of Edinburgh, School of Informatics, Edinburgh, EH8 9AB, UK

September 11, 2019

Archetypal Analysis

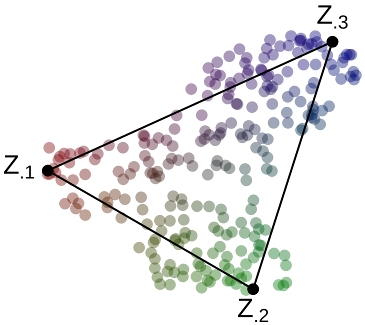
- *Archetypes* are prototypes, i.e., representative observations, that are ideal examples of a type
- Archetypes are interpretable since they relate to actual observations
- Archetypes are *extreme* in nature rather than *average*, for example, as in *medoids*



Archetypal Analysis

- *Archetypes* are prototypes, i.e., representative observations, that are ideal examples of a type
- Archetypes are interpretable since they relate to actual observations
- Archetypes are *extreme* in nature rather than *average*, for example, as in *medoids*

Can we find archetypes for observations that are not real valued?



Obs.	Favorite sports	Marital status	Gender	...
1	Tennis	Single	Female	
2	Golf	Divorced	Male	
3	Soccer	Divorced	Male	
4	Soccer	Divorced	Male	
5	Golf	Married	Female	
6	Tennis	Married	Female	
7	Tennis	Married	Male	
8	Golf	Single	Male	
9	Tennis	Single	Female	
10	Golf	Divorced	Male	
⋮				

Probabilistic Archetypal Analysis

Intuition

- Assume that the observations originate from a parametric probability distribution
- Perform archetypal analysis in the parameter space

Probabilistic principal component analysis

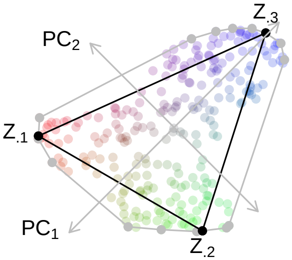
$$\mathbf{h}_n \sim \mathcal{N}(0, \mathbf{I}) \tag{1}$$

$$\mathbf{x}_n \sim \mathcal{N}(\mathbf{Z}\mathbf{h}_n, \sigma^2\mathbf{I}) \tag{2}$$

Probabilistic vertex component analysis

$$\mathbf{h}_n \sim \mathrm{Dir}(\mathbf{1}) \tag{3}$$

$$\mathbf{x}_n \sim \mathcal{N}(\mathbf{Z}\mathbf{h}_n, \sigma^2\mathbf{I}) \tag{4}$$



Probabilistic Archetypal Analysis

Intuition

- Assume that the observations originate from a parametric probability distribution
- Perform archetypal analysis in the parameter space

Probabilistic principal component analysis

$$\mathbf{h}_n \sim \mathcal{N}(0, \mathbf{I}) \quad (1)$$

$$\mathbf{x}_n \sim \mathcal{N}(\mathbf{Z}\mathbf{h}_n, \sigma^2\mathbf{I}) \quad (2)$$

Probabilistic vertex component analysis

$$\mathbf{h}_n \sim \text{Dir}(\mathbf{1}) \quad (3)$$

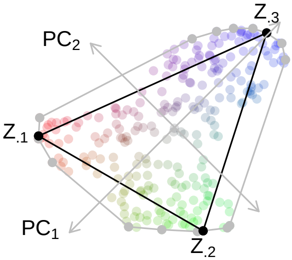
$$\mathbf{x}_n \sim \mathcal{N}(\mathbf{Z}\mathbf{h}_n, \sigma^2\mathbf{I}) \quad (4)$$

Probabilistic archetypal analysis: given $\mathbf{\Theta}$

$$\mathbf{w}_k \sim \text{Dir}(\mathbf{1}) \quad (5)$$

$$\mathbf{h}_n \sim \text{Dir}(\mathbf{1}) \quad (6)$$

$$\mathbf{x}_n \sim \mathcal{N}(\mathbf{\Theta}\mathbf{W}\mathbf{h}_n, \sigma^2\mathbf{I}) \quad (7)$$



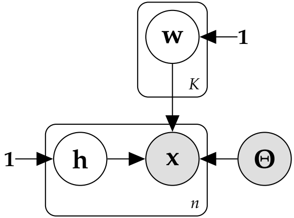
$$\mathbf{x}_n \sim \text{ExpFam}(\boldsymbol{\Theta}\mathbf{W}\mathbf{h}_n) \tag{8}$$

where

$$\text{ExpFam}(\mathbf{x}; \boldsymbol{\psi}) = h(\mathbf{x})g(\boldsymbol{\psi}) \exp(\boldsymbol{\eta}(\boldsymbol{\psi})^\top \mathbf{s}(\mathbf{x})) \tag{9}$$

is exponential family distributions, e.g.,

1. Normal for real valued observations, i.e., $[1.1, 0.3, 2.7]^\top$
   $\boldsymbol{\psi}$ is a vector of means, standard archetypal analysis



Probabilistic Archetypal Analysis

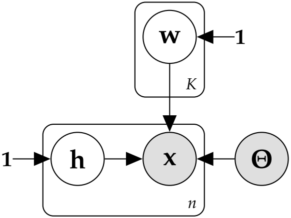
$$\mathbf{x}_n \sim \text{ExpFam}(\mathbf{\Theta W h}_n) \quad (8)$$

where

$$\text{ExpFam}(\mathbf{x};\boldsymbol{\psi}) = h(\mathbf{x})g(\boldsymbol{\psi})\exp(\boldsymbol{\eta}(\boldsymbol{\psi})^\top \mathbf{s}(\mathbf{x})) \quad (9)$$

is exponential family distributions, e.g.,

1. Normal for real valued observations, i.e., $[1.1, 0.3, 2.7]^\top$
   $\boldsymbol{\psi}$ is a vector of means, standard archetypal analysis
2. Poisson for count observations, i.e., $[1, 4, 2]^\top$
   $\boldsymbol{\psi}$ is a vector of rates



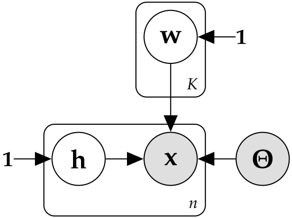
$$\mathbf{x}_n \sim \text{ExpFam}(\mathbf{\Theta W h}_n) \tag{8}$$

where

$$\text{ExpFam}(\mathbf{x}; \boldsymbol{\psi}) = h(\mathbf{x}) g(\boldsymbol{\psi}) \exp(\boldsymbol{\eta}(\boldsymbol{\psi})^\top \mathbf{s}(\mathbf{x})) \tag{9}$$

is exponential family distributions, e.g.,

1. Normal for real valued observations, i.e., $[1.1, 0.3, 2.7]^\top$
   $\boldsymbol{\psi}$ is a vector of means, standard archetypal analysis
2. Poisson for count observations, i.e., $[1, 4, 2]^\top$
   $\boldsymbol{\psi}$ is a vector of rates
3. Bernoulli for binary observations, i.e., $[\mathit{True}, \mathit{False}, \mathit{True}]^\top$
   $\boldsymbol{\psi}$ is a vector of probabilities



Probabilistic Archetypal Analysis

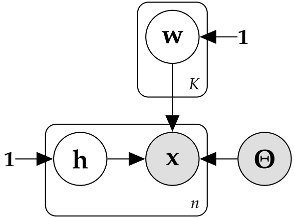
$$\mathbf{x}_n \sim \text{ExpFam}(\mathbf{\Theta W h}_n) \quad (8)$$

where

$$\text{ExpFam}(\mathbf{x}; \boldsymbol{\psi}) = h(\mathbf{x})g(\boldsymbol{\psi}) \exp(\boldsymbol{\eta}(\boldsymbol{\psi})^\top \mathbf{s}(\mathbf{x})) \quad (9)$$

is exponential family distributions, e.g.,

1. Normal for real valued observations, i.e., $[1.1, 0.3, 2.7]^\top$
   $\boldsymbol{\psi}$ is a vector of means, standard archetypal analysis
2. Poisson for count observations, i.e., $[1, 4, 2]^\top$
   $\boldsymbol{\psi}$ is a vector of rates
3. Bernoulli for binary observations, i.e., $[\mathit{True}, \mathit{False}, \mathit{True}]^\top$
   $\boldsymbol{\psi}$ is a vector of probabilities
4. Multinomial for term-frequency values, i.e., $[60, 51, 42]^\top$
   $\boldsymbol{\psi}$ is a stochastic vector (number of trials is known)



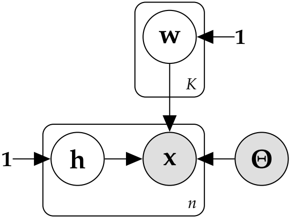
$$\mathbf{x}_n \sim \text{ExpFam}(\boldsymbol{\Theta}\mathbf{W}\mathbf{h}_n) \quad (8)$$

where

$$\text{ExpFam}(\mathbf{x}; \boldsymbol{\psi}) = h(\mathbf{x})g(\boldsymbol{\psi})\exp(\boldsymbol{\eta}(\boldsymbol{\psi})^\top \mathbf{s}(\mathbf{x})) \quad (9)$$

is exponential family distributions, e.g.,

1. Normal for real valued observations, i.e., $[1.1, 0.3, 2.7]^\top$
   $\boldsymbol{\psi}$ is a vector of means, standard archetypal analysis
2. Poisson for count observations, i.e., $[1, 4, 2]^\top$
   $\boldsymbol{\psi}$ is a vector of rates
3. Bernoulli for binary observations, i.e., $[\mathit{True}, \mathit{False}, \mathit{True}]^\top$
   $\boldsymbol{\psi}$ is a vector of probabilities
4. Multinomial for term-frequency values, i.e., $[60, 51, 42]^\top$
   $\boldsymbol{\psi}$ is a stochastic vector (number of trials is known)
5. Multinomials for categorical observations, i.e., $[\mathit{Tennis}, \mathit{Single}, \mathit{Female}]^\top$
   $\boldsymbol{\psi}$ is a stochastic 'matrix' (number of trials is known)



Probabilistic Archetypal Analysis

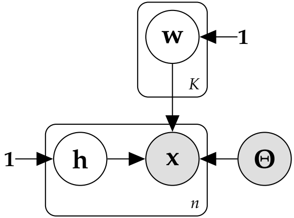
$$\mathbf{x}_n \sim \text{ExpFam}(\mathbf{\Theta}\mathbf{W}\mathbf{h}_n) \tag{8}$$

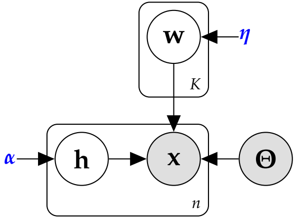
where

$$\text{ExpFam}(\mathbf{x};\boldsymbol{\psi}) = h(\mathbf{x})g(\boldsymbol{\psi})\exp(\boldsymbol{\eta}(\boldsymbol{\psi})^\top \mathbf{s}(\mathbf{x})) \tag{9}$$

is exponential family distributions, e.g.,

1. Normal for real valued observations, i.e., $[1.1, 0.3, 2.7]^\top$
   $\boldsymbol{\psi}$ is a vector of means, standard archetypal analysis
2. Poisson for count observations, i.e., $[1, 4, 2]^\top$
   $\boldsymbol{\psi}$ is a vector of rates
3. Bernoulli for binary observations, i.e., $[\mathit{True}, \mathit{False}, \mathit{True}]^\top$
   $\boldsymbol{\psi}$ is a vector of probabilities
4. Multinomial for term-frequency values, i.e., $[60, 51, 42]^\top$
   $\boldsymbol{\psi}$ is a stochastic vector (number of trials is known)
5. Multinomials for categorical observations, i.e., $[\mathit{Tennis}, \mathit{Single}, \mathit{Female}]^\top$
   $\boldsymbol{\psi}$ is a stochastic 'matrix' (number of trials is known)





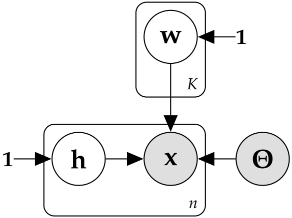
$$\mathbf{x}_n \sim \text{ExpFam}(\mathbf{\Theta}\mathbf{W}\mathbf{h}_n) \quad (8)$$

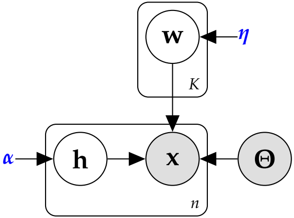
where

$$\text{ExpFam}(\mathbf{x};\boldsymbol{\psi}) = h(\mathbf{x})g(\boldsymbol{\psi})\exp(\boldsymbol{\eta}(\boldsymbol{\psi})^\top \mathbf{s}(\mathbf{x})) \quad (9)$$

is exponential family distributions, e.g.,

1. Normal for real valued observations, i.e., $[1.1, 0.3, 2.7]^\top$
   $\boldsymbol{\psi}$ is a vector of means, standard archetypal analysis
2. Poisson for count observations, i.e., $[1, 4, 2]^\top$
   $\boldsymbol{\psi}$ is a vector of rates
3. Bernoulli for binary observations, i.e., $[\mathit{True}, \mathit{False}, \mathit{True}]^\top$
   $\boldsymbol{\psi}$ is a vector of probabilities
4. Multinomial for term-frequency values, i.e., $[60, 51, 42]^\top$
   $\boldsymbol{\psi}$ is a stochastic vector (number of trials is known)
5. Multinomials for categorical observations, i.e., $[\mathit{Tennis}, \mathit{Single}, \mathit{Female}]^\top$
   $\boldsymbol{\psi}$ is a stochastic 'matrix' (number of trials is known)





We consider $\boldsymbol{\theta}_n$ to be the *maximum likelihood point estimate* from observation $\mathbf{x}_n$

Probabilistic Archetypal Analysis

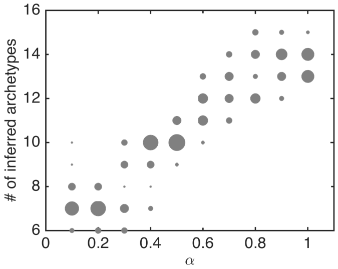
- PAA solves

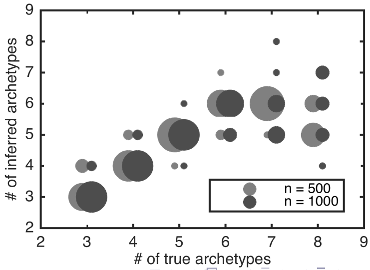
$$\underset{\mathbf{W},\mathbf{H}\geq 0}{\arg\min} -\mathbb{LL}(\mathbf{X}|\mathbf{W},\mathbf{H},\boldsymbol{\Theta}) \text{ such that } \mathbf{1W}=\mathbf{1},\mathbf{1H}=\mathbf{1}. \tag{10}$$

under suitable observation model compared to AA that solves

$$\min_{\mathbf{W},\mathbf{H}\geq 0} ||\mathbf{X}-\mathbf{XWH}||_F^2 \text{ such that } \mathbf{1W}=\mathbf{1},\mathbf{1H}=\mathbf{1} \tag{11}$$

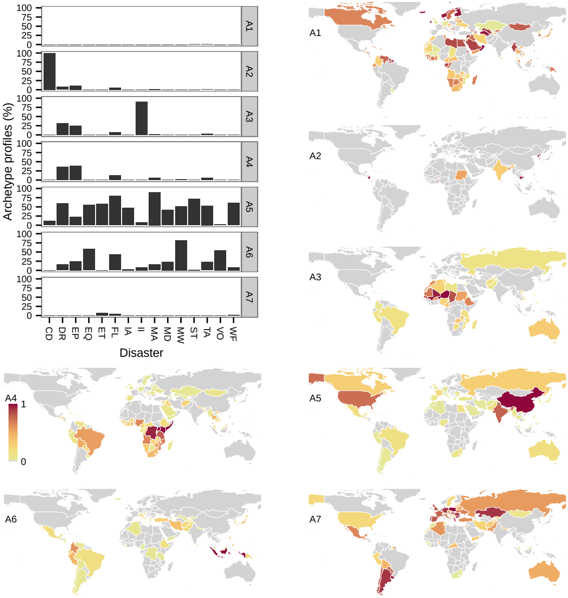
- This optimization problem can be solved using majorization-minimization for Poisson likelihood, and expectation maximization and variational Bayes' for multinomial observation model and the latter takes advantage of conjugacy
- K can be chosen using 'elbow criterion' or setting prior over $\boldsymbol{\alpha}$ over $\mathbf{h}$





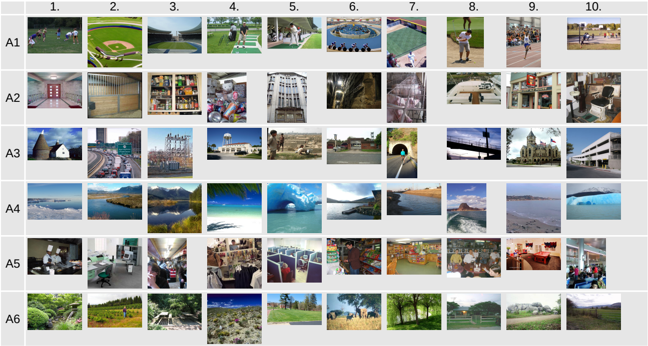
Disasters worldwide 1900-2008

1. complex disasters (CD),
2. drought (DR),
3. earthquake (EQ),
4. epidemic (EP),
5. extreme temperature (ET),
6. flood (FL),
7. industrial accident (IA),
8. insect infestation (II),
9. mass movement dry (MD),
10. mass movement wet (MW),
11. miscellaneous accident (MA),
12. storm (ST),
13. transport accident (TA),
14. volcano (VO), and
15. wildfire (WF).

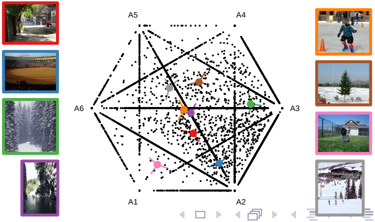


SUN attribute

14340 images with 102 attributes with 4 annotations each converted to binary



A1 outdoor sport activities
A2 abstract images
A3 buildings and constructions
A4 natural blue and horizon
A5 enclosed area
A6 natural green and trees



Big5 Personality

19718 responses, 50 Likert-scale rated questions converted to {agree, disagree, neutral}

Summary

- PAA finds extreme representations for non-real-valued observations
- PAA also helps in choosing an appropriate number of archetypes
- PAA provides more distinguishable representation compared to clustering

Thank you for your patience!

- Sohan Seth and Manuel J. A. Eugster. Probabilistic archetypal analysis. *Machine Learning*, 102(1):85–113, January 2016.
- Sohan Seth and Manuel J. A. Eugster. Archetypal Analysis for Nominal Observations. *IEEE transactions on pattern analysis and machine intelligence*, 38(5):849–861, May 2016.
- `http://aalab.github.io`