$H_2O$.ai

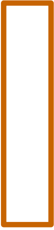
# GLRM Overview

- GLRM is an extension of well-known matrix factorization methods such as Principal Component Analysis (PCA).
- Unlike PCA which is limited to numerical data, GLRM can also handle categorical, ordinal and Boolean data.
- **Given**: Data table *A* with *m* rows and *n* columns
- **Find**: Compressed representation as numeric tables *X* and *Y* where *k* is a small user-specified number

- *Y* = archetypal features created from columns of *A*
- *X* = row of A in reduced feature space
- GLRM can approximately reconstruct *A* from product *XY*

230

$$
m\left\{\overbrace{\begin{bmatrix} & & \\ & A & \\ & & \end{bmatrix}}^{n}\right. \approx m\left\{\overbrace{\begin{bmatrix} \\ X \\ \\ \end{bmatrix}}^{k}\right. \left.\overbrace{\begin{bmatrix} & Y & \end{bmatrix}}^{n}\right\}k
$$



$\approx$

+

# Memory Reduction / Saving

# GLRM Loss Functions

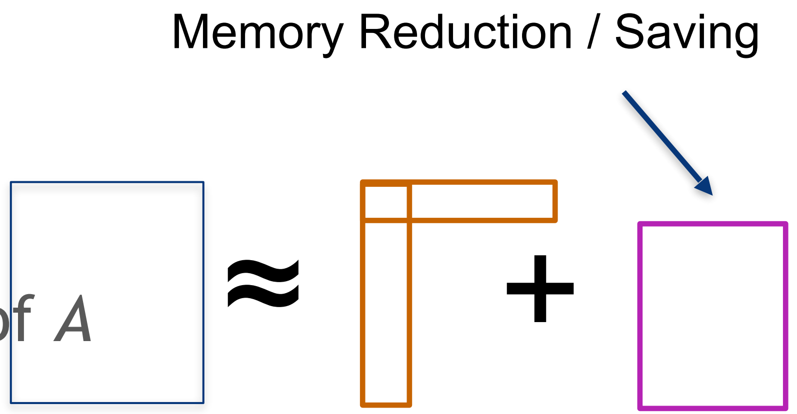
Each column can use different loss function

| Loss | Principal Components Analysis |
| --- | --- |
| Quadratic | PCA |
| Absolute | Robust PCA |
| Huber | Huber PCA (Hybrid of Quadratic and Robust) |
| Poisson | Poisson PCA |
| Hinge | Boolean PCA |
| Logistic | Logistic PCA |
| Periodic | Periodic PCA |
| Categorica l | Categorical PCA |
| Ordinal | Ordinal PCA |

# GLRM Overview

- GLRM is an extension of well-known matrix factorization methods such as Principal Component Analysis (PCA).
- Unlike PCA which is limited to numerical data, GLRM can also handle categorical, ordinal and Boolean data.
- **Given**: Data table *A* with *m* rows and *n* columns
- **Find**: Compressed representation as numeric tables *X* and *Y* where *k* is a small user-specified number

$$m\left\{\left[\overbrace{\quad A \quad}^{n}\right]\right. \approx m\left\{\left[\overbrace{X}^{k}\right]\right. \left.\left[\overbrace{\quad Y \quad}^{n}\right]\right\}k$$




- *Y* = archetypal features created from columns of *A*
- *X* = row of A in reduced feature space
- GLRM can approximately reconstruct *A* from product *XY*