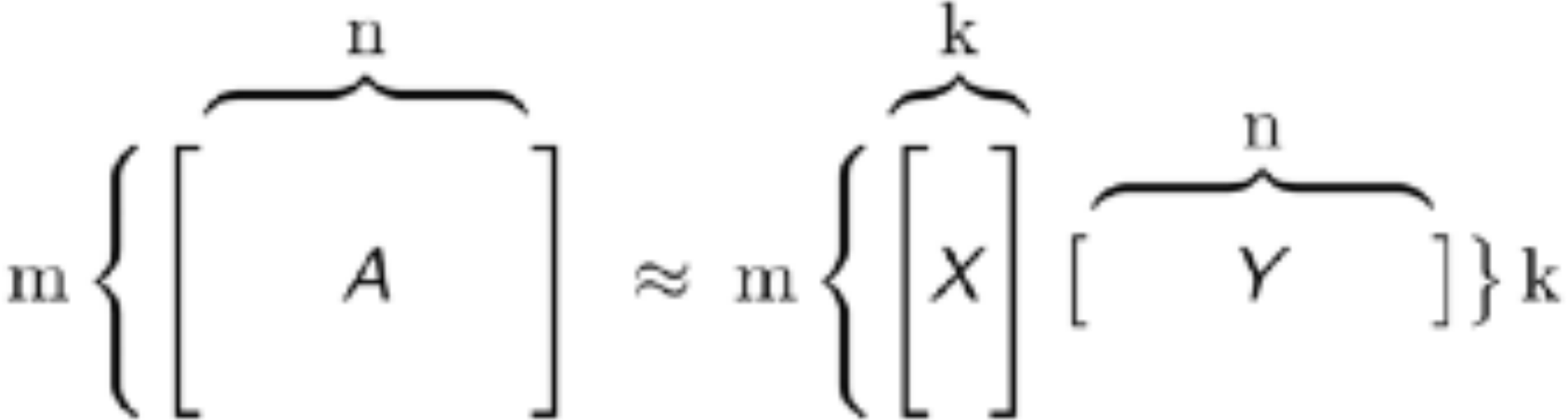
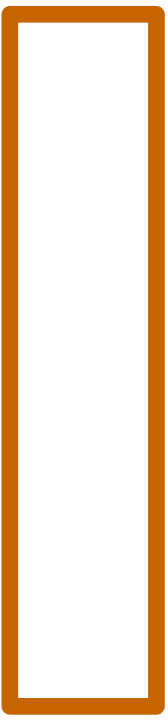

H2O.ai

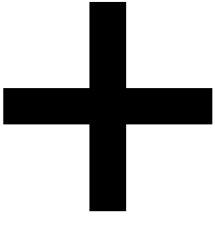
Generalized Low-Rank Models

- 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

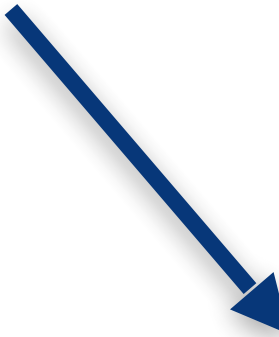
- *Y* = archetypal features created from columns of *A*
- *X* = row of A in reduced feature space
- GLRM can approximately reconstruct *A* from product *XY*


n
m
A
≈
k
m
X
n
Y
k





Memory Reduction / Saving



H2O Generalized Low-Rank Models

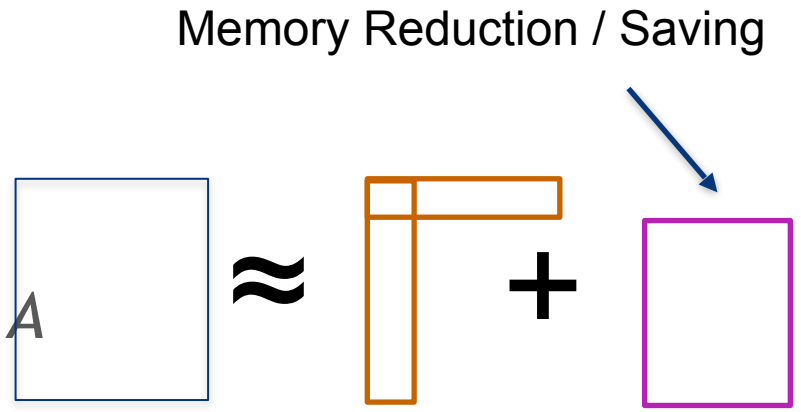
```
h2o.glrm(training_frame, cols = NULL, model_id = NULL, validation_frame = NULL,
         ignore_const_cols = TRUE, score_each_iteration = FALSE, loading_name = NULL,
         transform = c("NONE", "STANDARDIZE", "NORMALIZE", "DEMEAN", "DESCALE"), k = 1,
         loss = c("Quadratic", "Absolute", "Huber", "Poisson", "Hinge", "Logistic", "Periodic"),
         loss_by_col = c("Quadratic", "Absolute", "Huber", "Poisson", "Hinge", "Logistic",
         "Periodic", "Categorical", "Ordinal"), loss_by_col_idx = NULL,
         multi_loss = c("Categorical", "Ordinal"), period = 1,
         regularization_x = c("None", "Quadratic", "L2", "L1", "NonNegative", "OneSparse",
         "UnitOneSparse", "Simplex"), regularization_y = c("None", "Quadratic", "L2", "L1",
         "NonNegative", "OneSparse", "UnitOneSparse", "Simplex"), gamma_x = 0, gamma_y = 0,
         max_iterations = 1000, max_updates = 2000, init_step_size = 1,
         min_step_size = 1e-04, seed = -1, init = c("Random", "SVD", "PlusPlus", "User"),
         svd_method = c("GramSVD", "Power", "Randomized"), user_y = NULL, user_x = NULL,
         expand_user_y = TRUE, impute_original = FALSE, recover_svd = FALSE, max_runtime_secs = 0)
```

```
from h2o.estimators.glrm import H2OGeneralizedLowRankEstimator
model = H2OGeneralizedLowRankEstimator(…)
model.train(x = x, training_frame = data)
```

Generalized Low-Rank Models

- 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

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



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