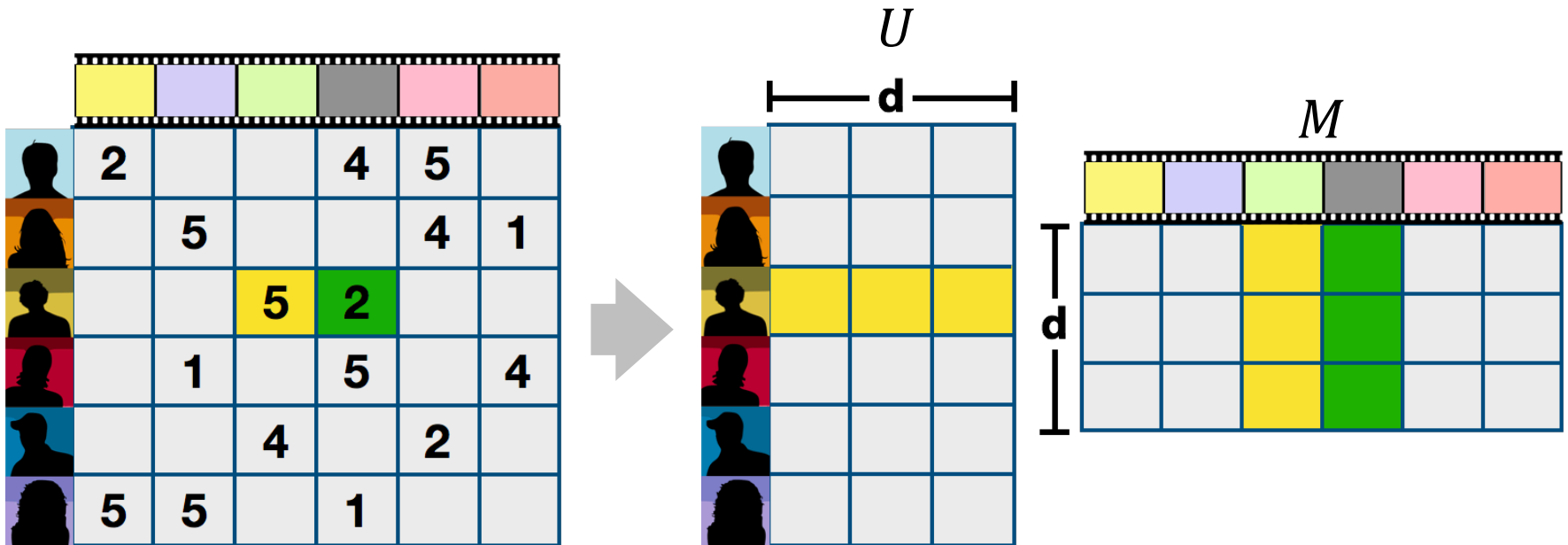


Matrix Factorizations for Feature Extraction

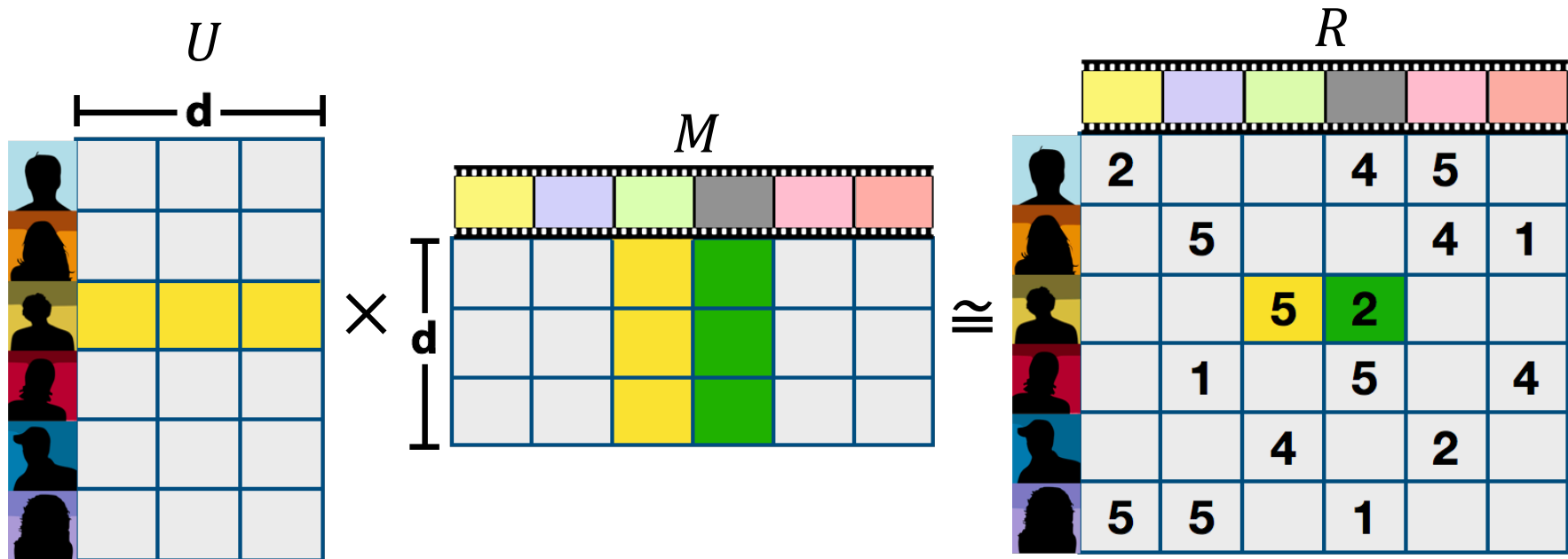
Example

	2			4	5	
		5			4	1
			5	2		
		1		5		4
			4		2	
	5	5		1		

Example



What is Matrix Factorization?



Documents\words example

	cat	is	dog	animal	nature
Cat is animal	1	0	0	1	0
Dog is animal	0	1	1	1	0
Cat is dog	1	1	1	0	0

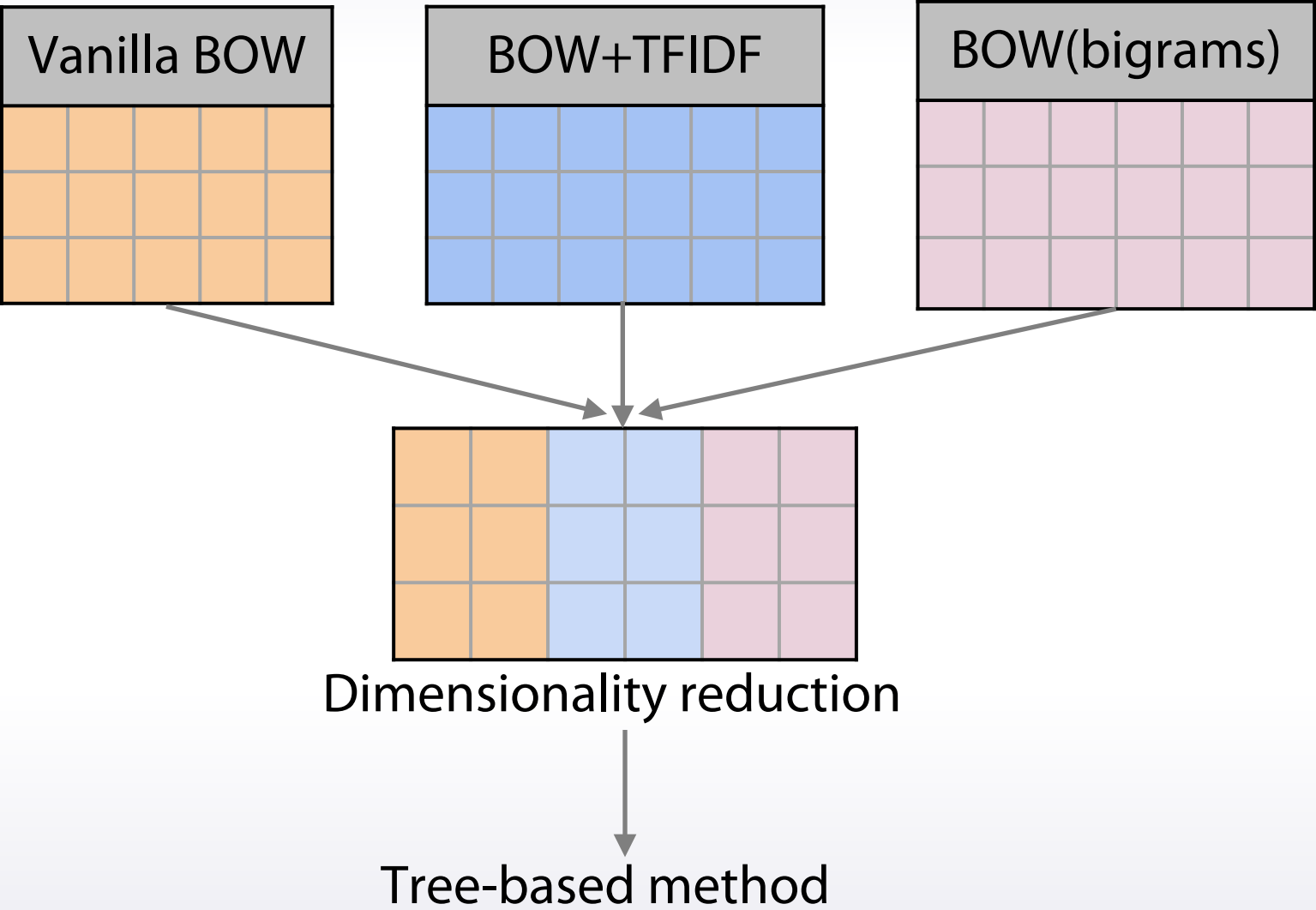
Documents\words example

	cat	is	dog	animal	nature
Cat is animal	1	0	0	1	0
Dog is animal	0	1	1	1	0
Cat is dog	1	1	1	0	0

Cat is animal			
Dog is animal			
Cat is dog			

cat			
is			
dog			
animal			
nature			

Example of feature fusion



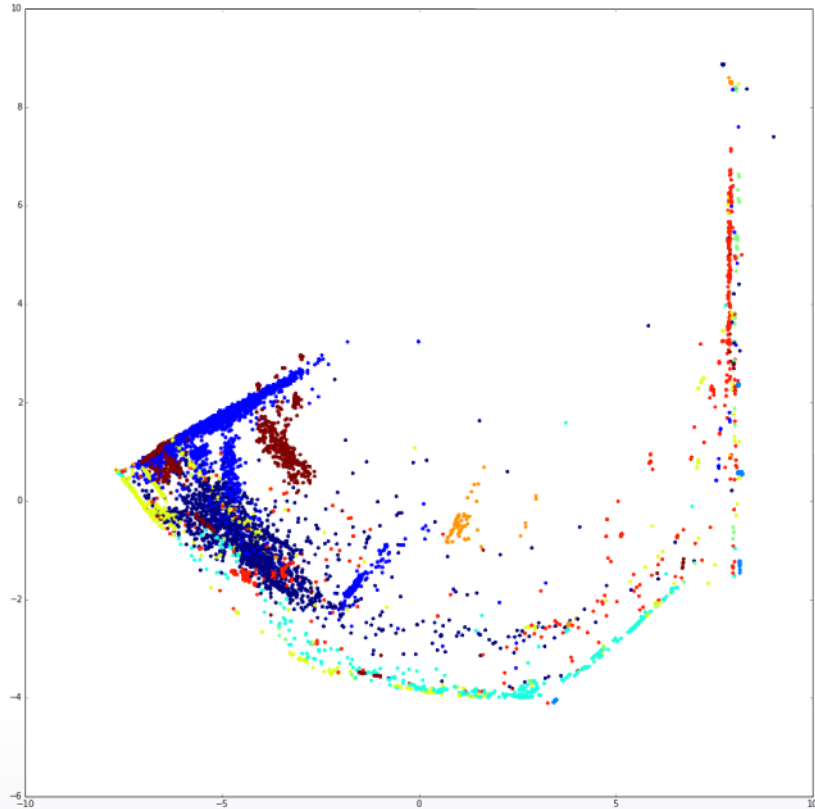
Notes about Matrix Factorization

- Can be apply only for some columns
- Can provide additional diversity
 - Good for ensembles
- It is a **lossy** transformation. Its' efficiency depends on:
 - Particular task
 - Number of latent factors
 - Usually 5-100

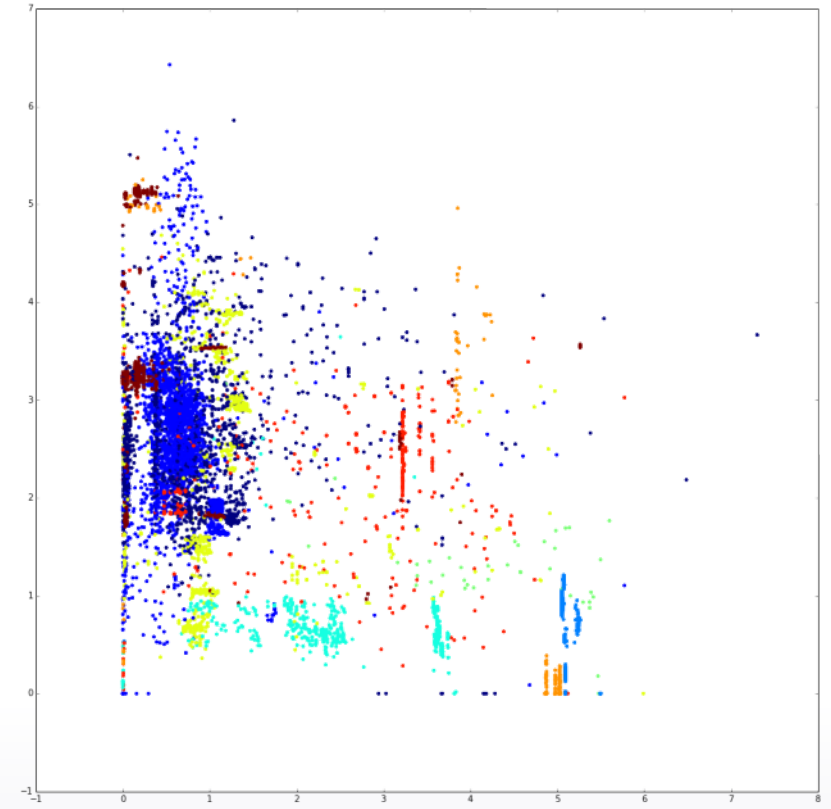
Implementation

- Several MF methods you can find in sklearn
- SVD and PCA
 - Standard tools for Matrix Factorization
- TruncatedSVD
 - Works with sparse matrices
- Non-negative Matrix Factorization (NMF)
 - Ensures that all latent factors are non-negative
 - Good for counts-like data

NMF for tree-based methods



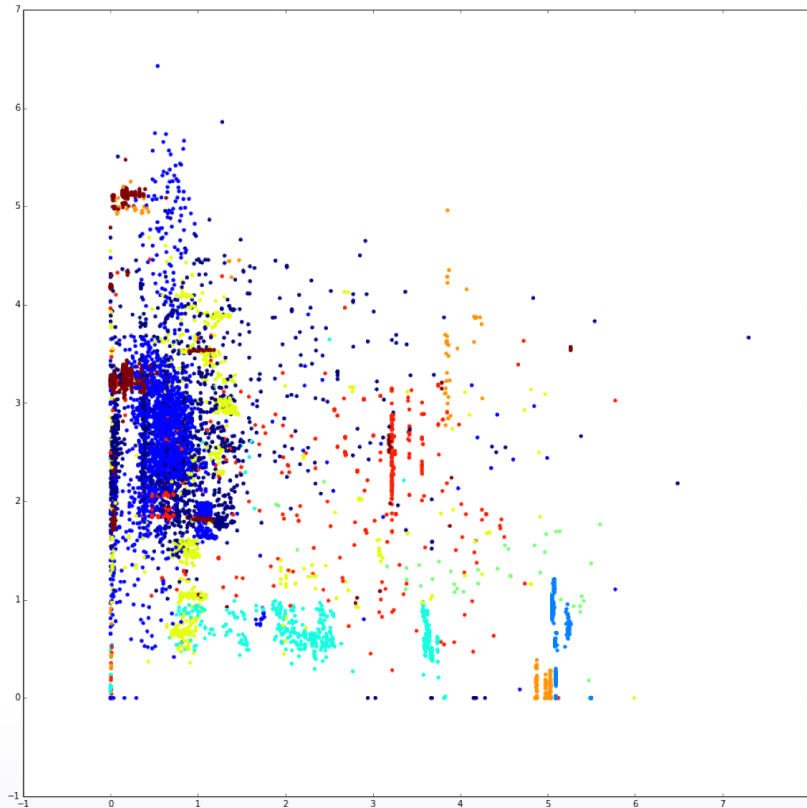
PCA



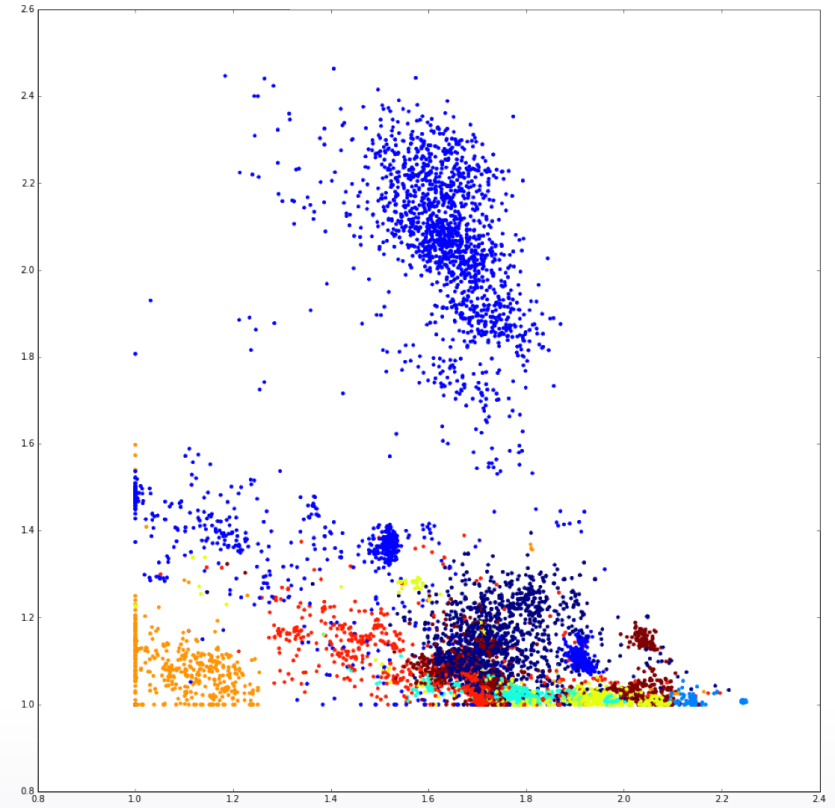
NMF

Notes about MF

Matrix factorization is similar in spirit to linear models.
You can use the same transformation tricks.



NMF(X)



NMF($\log(X+1)$)

Gochas

Wrong way:

```
pca = PCA(n_components=5)
X_train_pca = pca.fit_transform(X_train)
X_test_pca = pca.fit_transform(X_test)
```

Right way:

```
X_all = np.concatenate([X_train, X_test])
pca.fit(X_all)
X_train_pca = pca.transform(X_train)
X_test_pca = pca.transform(X_test)
```

Conclusion

- Matrix Factorization is a very general approach for dimensionality reduction and feature extraction

Conclusion

- Matrix Factorization is a very general approach for dimensionality reduction and feature extraction
- It can be applied for transforming categorical features into real-valued

Conclusion

- Matrix Factorization is a very general approach for dimensionality reduction and feature extraction
- It can be applied for transforming categorical features into real-valued
- Many of tricks suitable for linear models can be useful for MF