Collaborative Filtering : Recommendation System through different methods

Dakhlaoui Rayane

Khosrowshahi Sacha
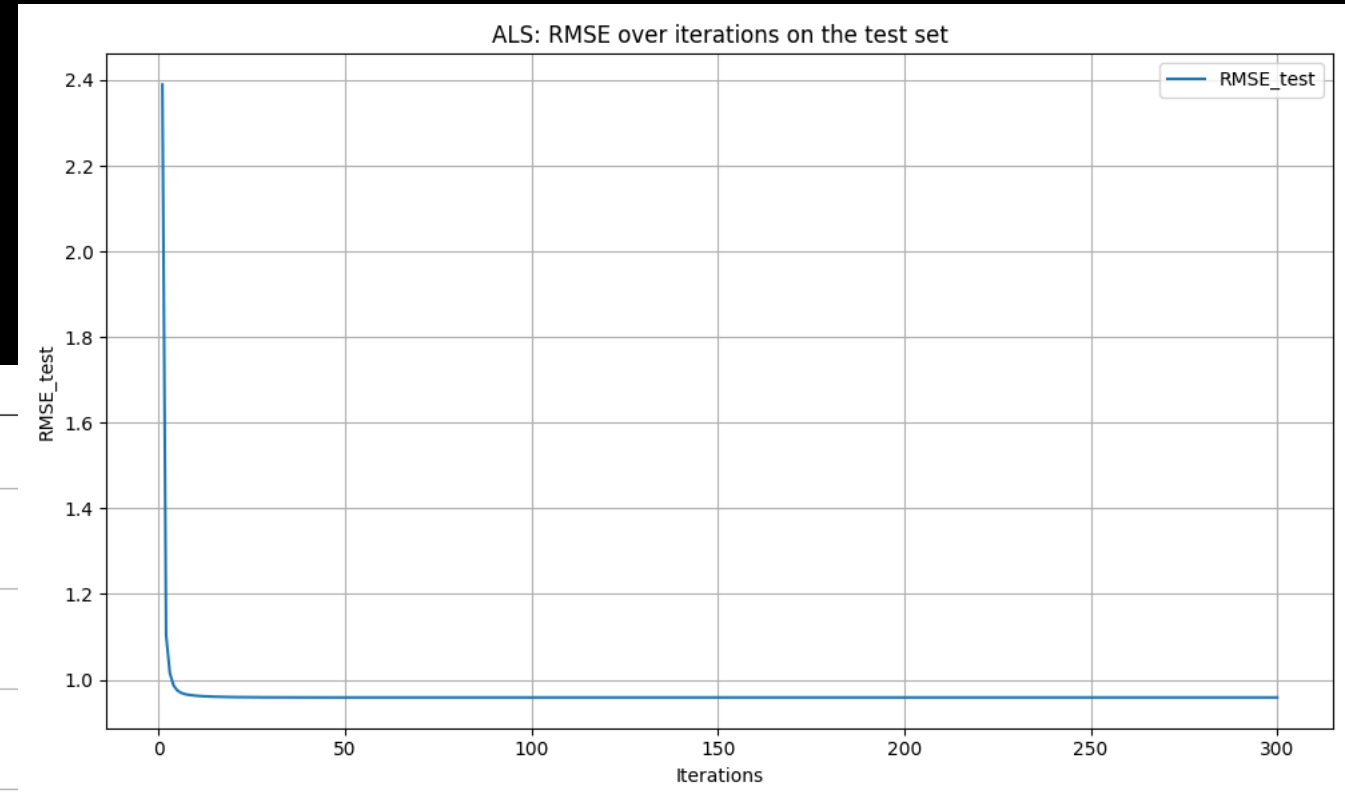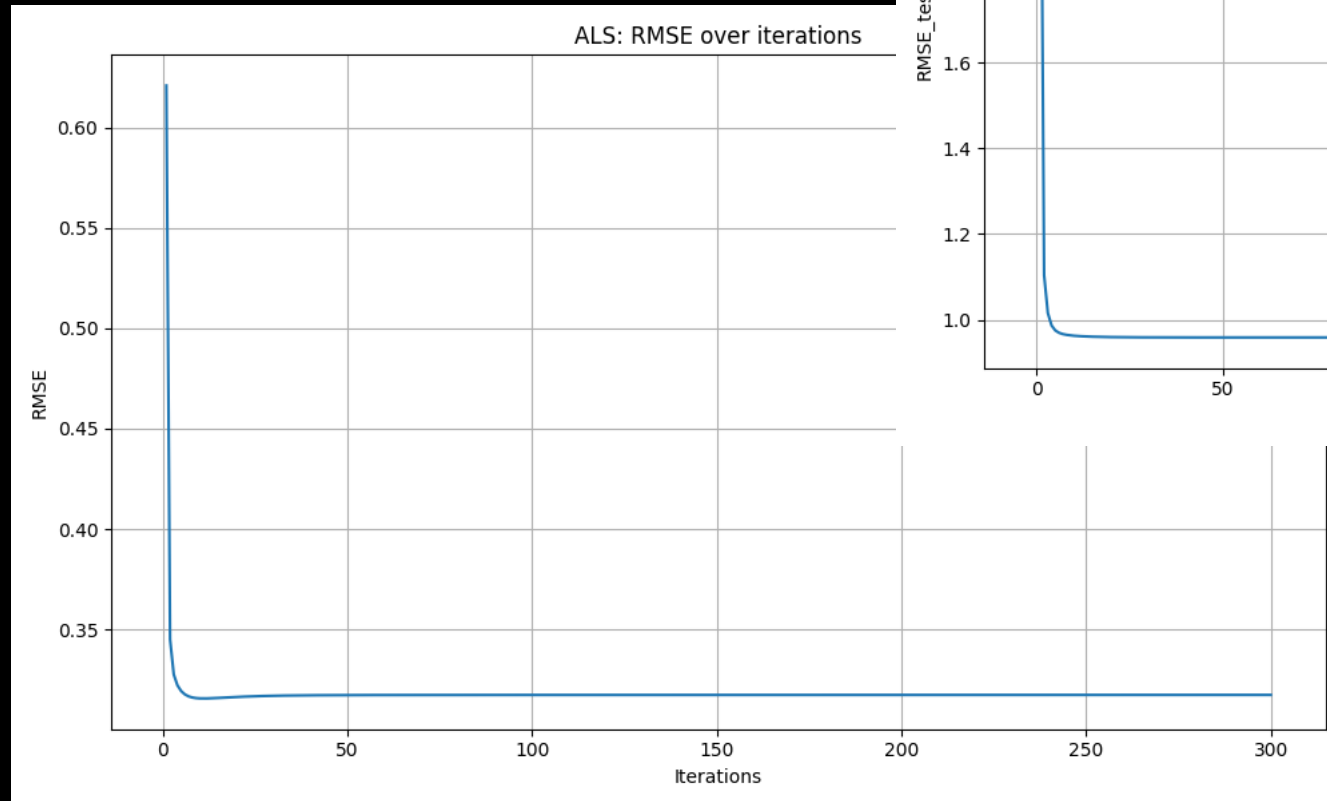
Rouillé Nathan

Index

- Matrix Factorization : ALS and Gradient descent

- Kernel MF

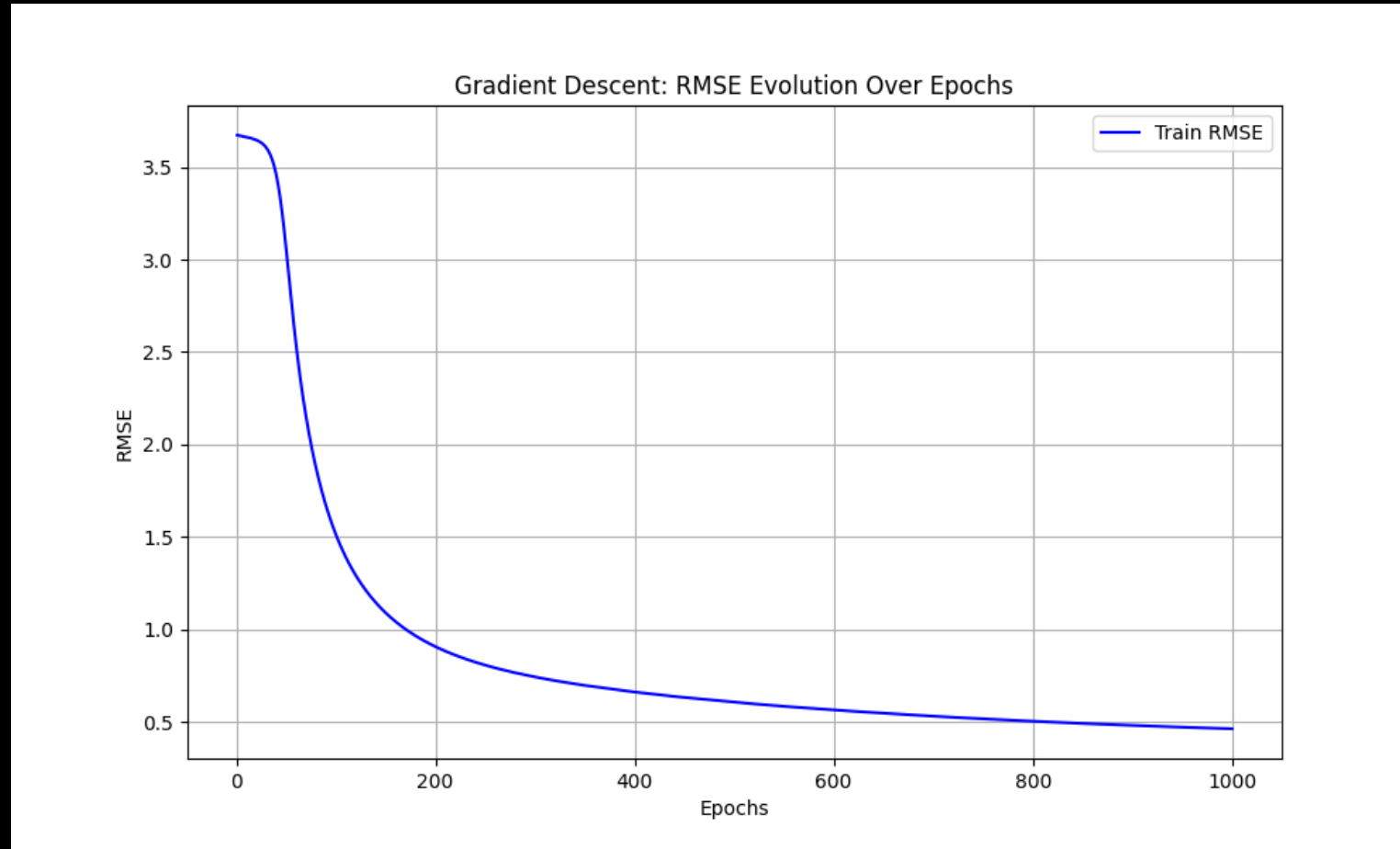- PCA : Method, Interpretation and improvements
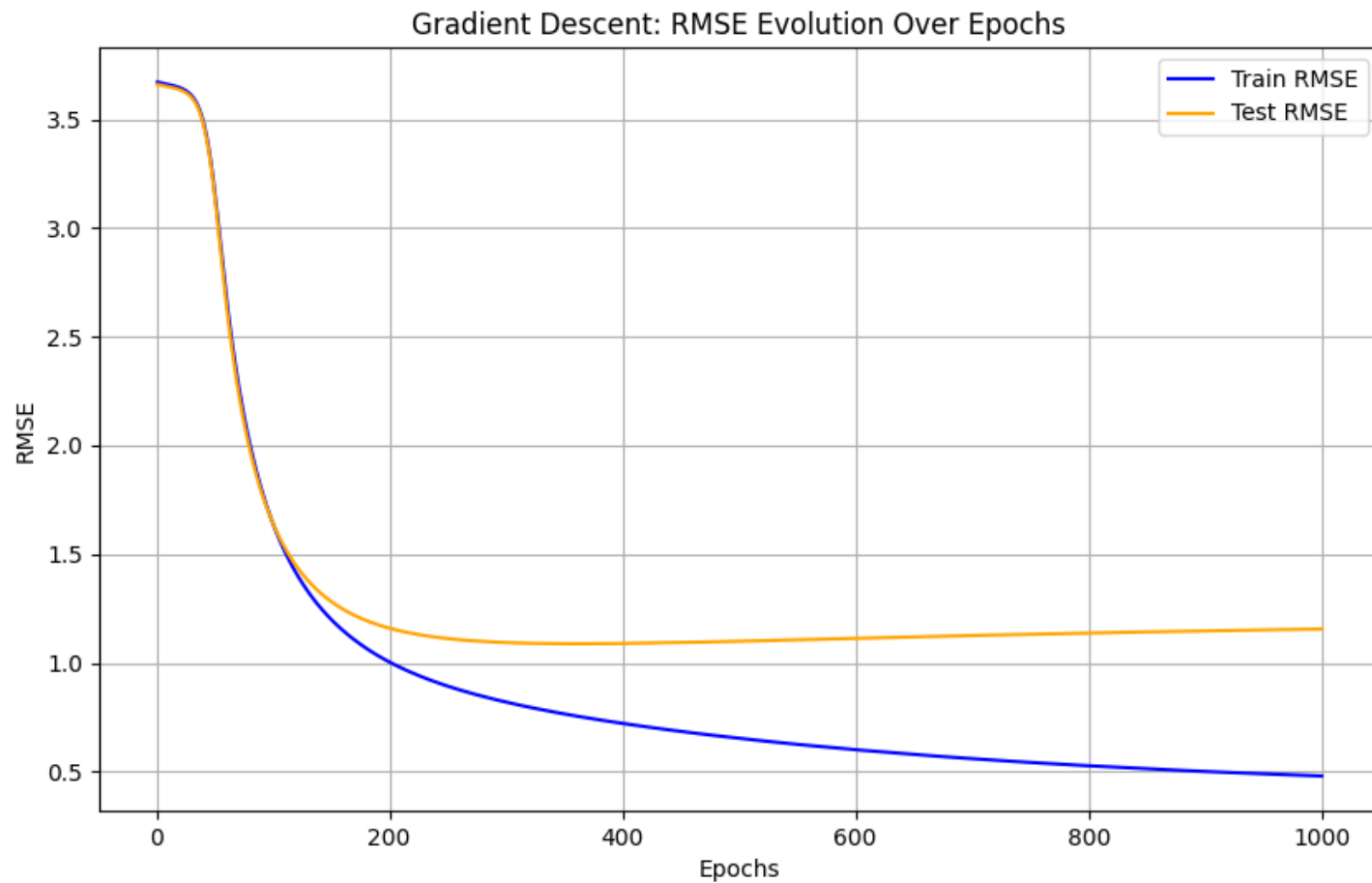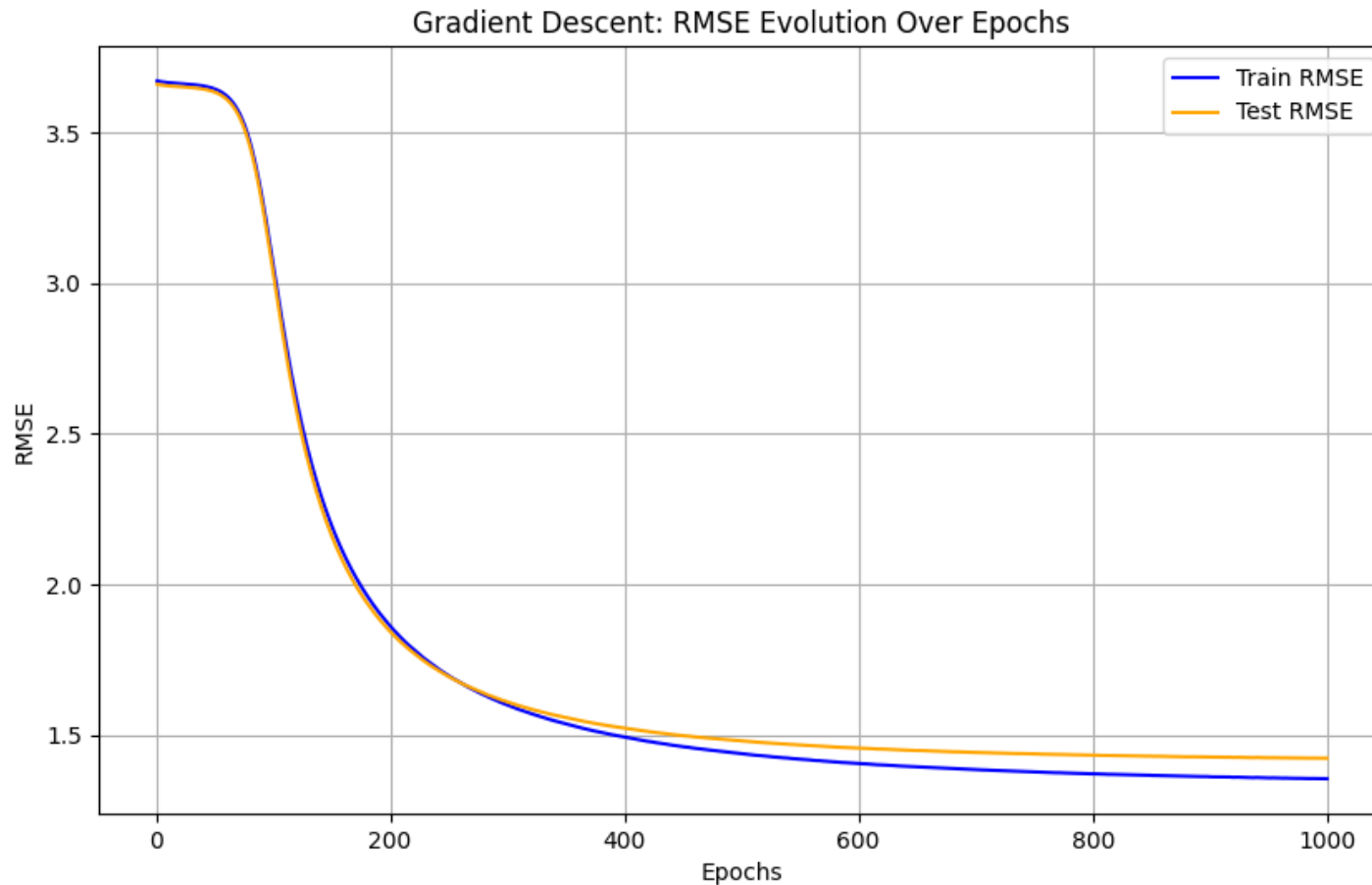
ALS analysis

Training

Testing

SGD – A better alternative that needs more tuning

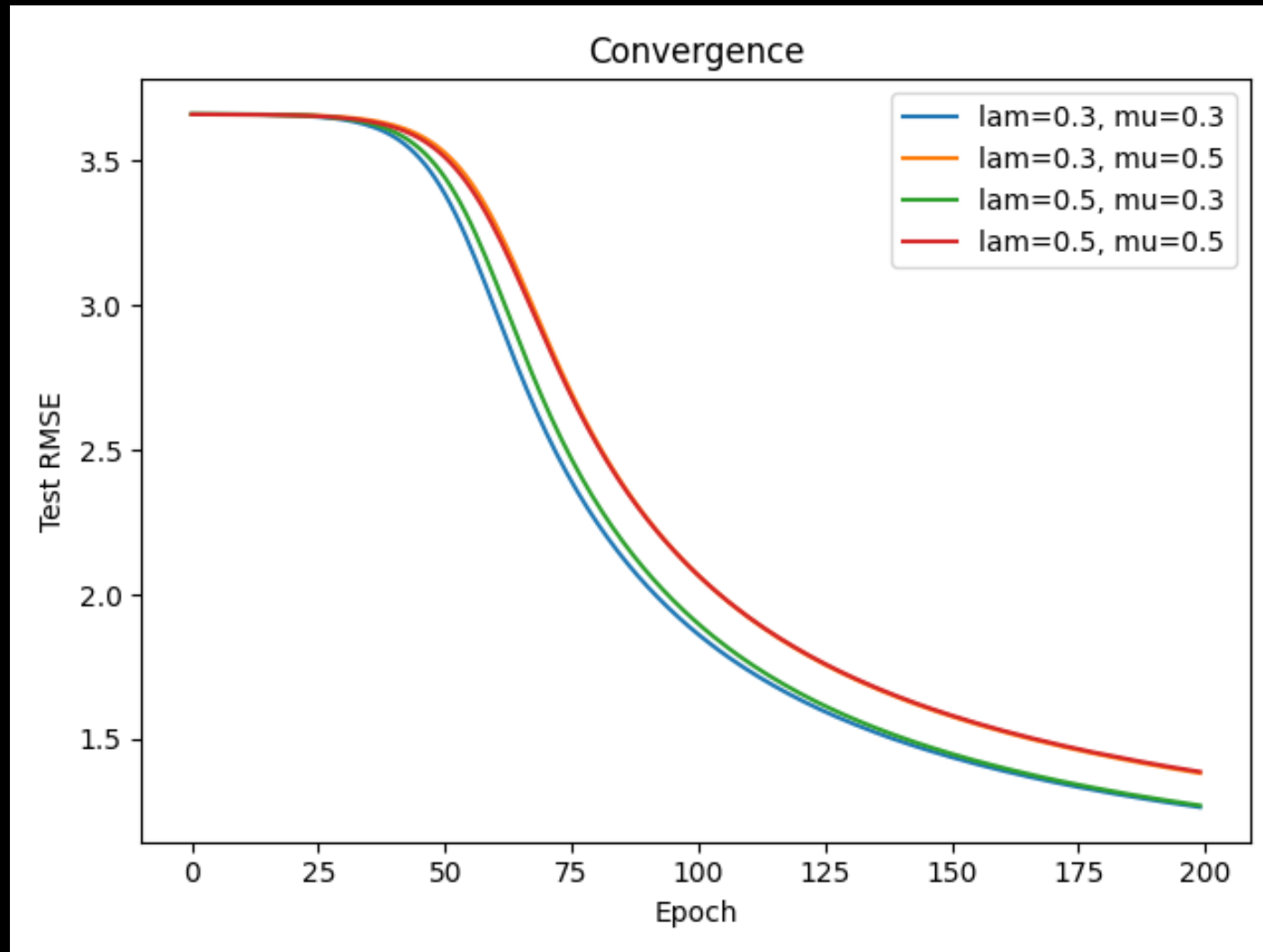# Overfitting

Gradient Descent: RMSE Evolution Over Epochs

Adding strong regularization
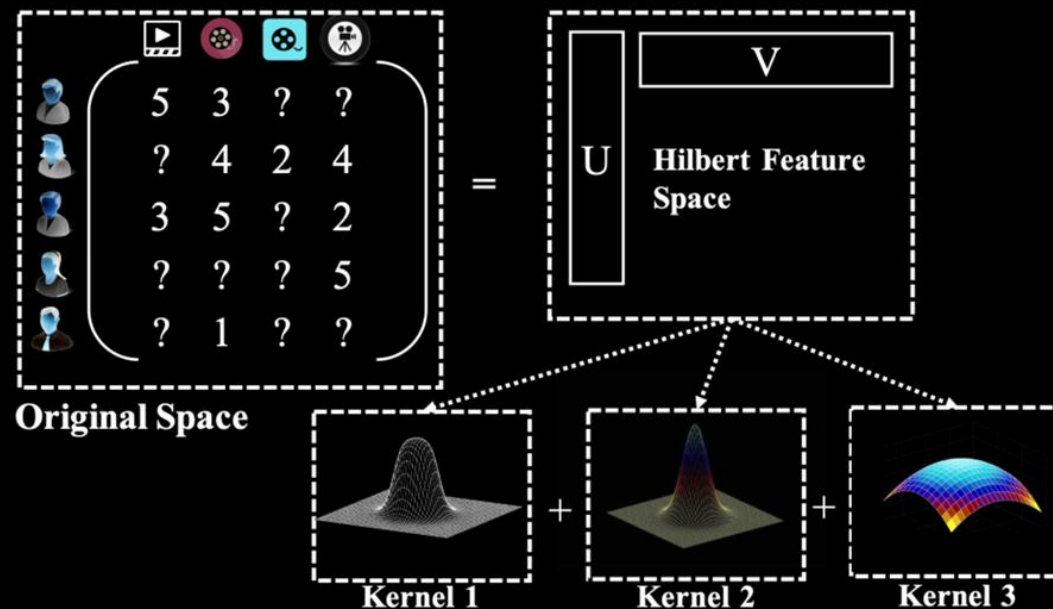
Grid search for best hyperparameters

Takeaways

- ALS converges fast and monotonically on MSE; early stop near 10–30 iterations.

- Best test RMSE in our sweep ≈0.97 (λ=0.30, μ=0.30, k=80).

- SGD reaches better quality with careful LR/regularization, but needs more tuning/time.

Kernel Matrix Factorization

- Why kernel in our case ?

$$R_{ui} \approx U_u^\top I_i \qquad \Longrightarrow \qquad R_{ui} = \kappa(U_u, I_i) = \langle \phi(U_u), \phi(I_i) \rangle$$
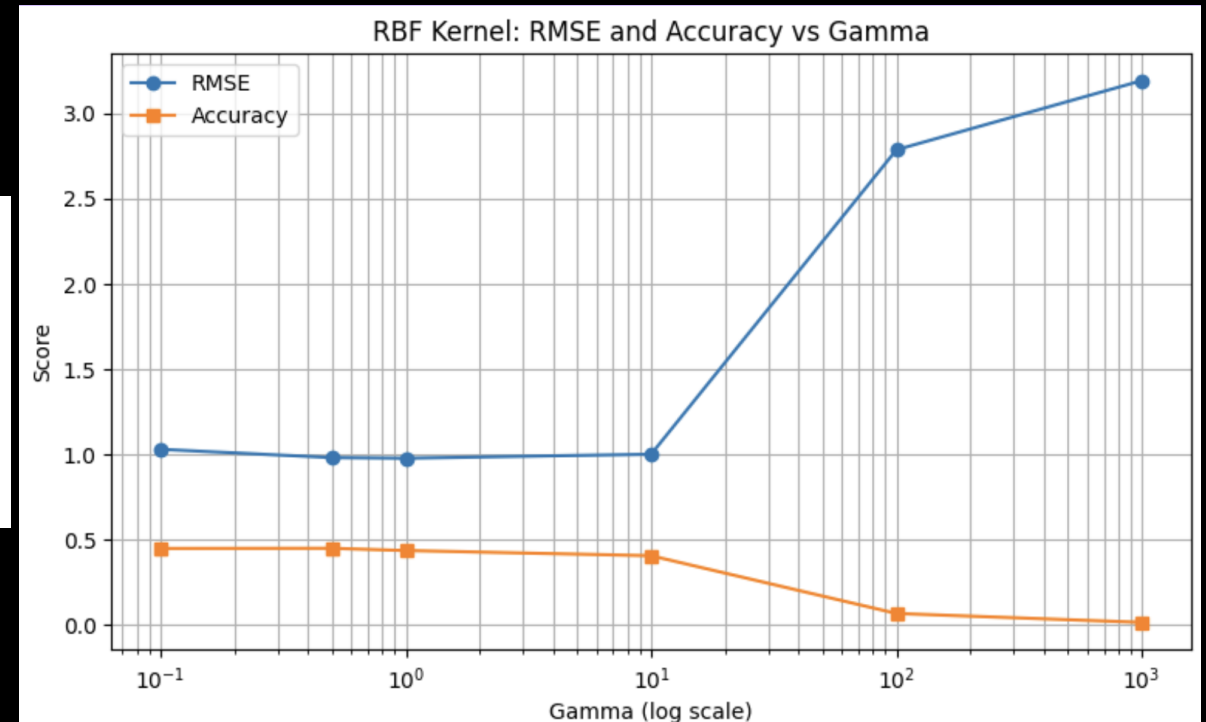

Kernelized Matrix Factorization for Collaborative Filtering (Liu et al, 2016)

- Predicted rating = (Global mean) + (user and item bias) + (interaction between latent factors)

Experiment for 3 different kernels :

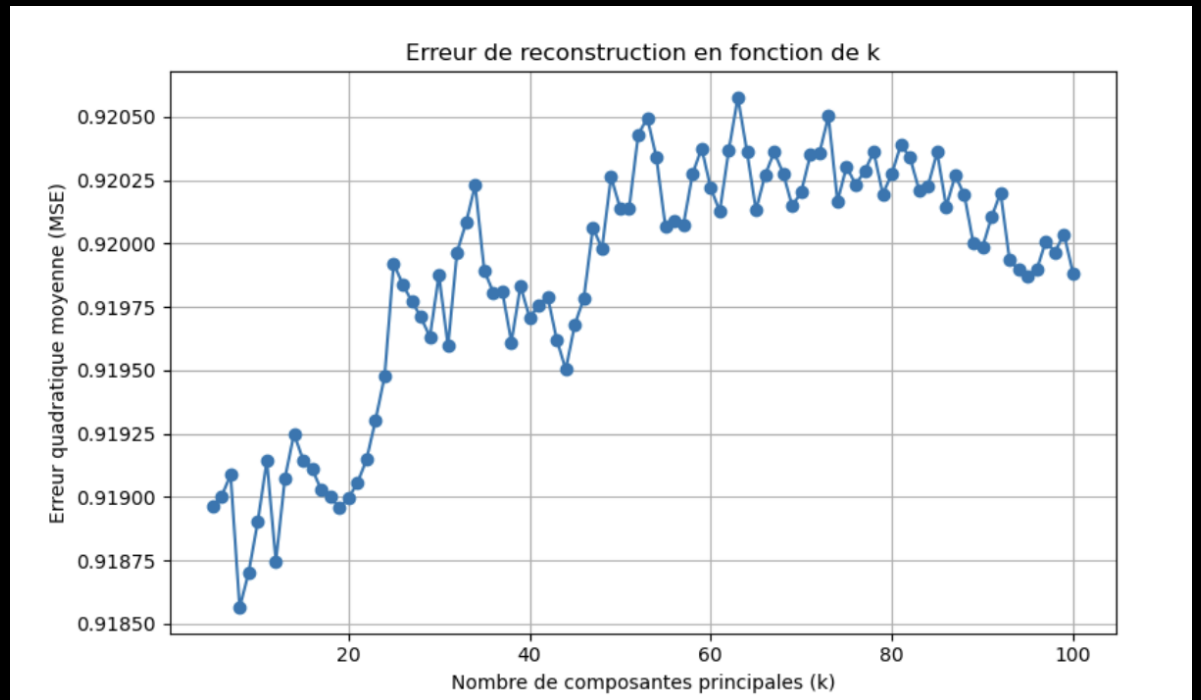|	RBF Kernel	Sigmoid Kernel	Linear Kernel
test RMSE	0.9921	1.0615	0.9139
me for 600 epoch	5.29 sec	4.29 sec	1.46 sec

RBF Kernel: RMSE and Accuracy vs Gamma

PCA

- Algo : partial covariance + EVD + round

- Hyperparameter optimization :

We have the best results for k=8
dimensions

- Performance test :

- RMSE = 0.92

- Accuracy = 0.24

Erreur de reconstruction en fonction de k

Interpretation of principal components

- Most correlated genres to each component:
- Component 1 : Comedy
- Component 2 : Romance
- Component 3 : Fantasy
- Component 4 : Drama
- Component 5 : Children
- Component 6 : Drama/Horror
- Component 7 : Film-Noir
- Component 8 : Crime

PCA Improvements

- Iterative PCA : EM algorithm

  - E-step : imputation missing data

  - M-step : compute PCA

  - Problem : hard rounding (smooth rounding linear/sigmoïd)

  - Performance test : RMSE 0.93 / Accuracy 0.24 (both hard & smooth rounding)

- KPCA : Sanguinetti & Lawrence paper and implementation in MATLAB

  - Problem : computationally expensive

- Hybride PCA + MF : use PCA as initialisation of MF

Methods for the upcoming week