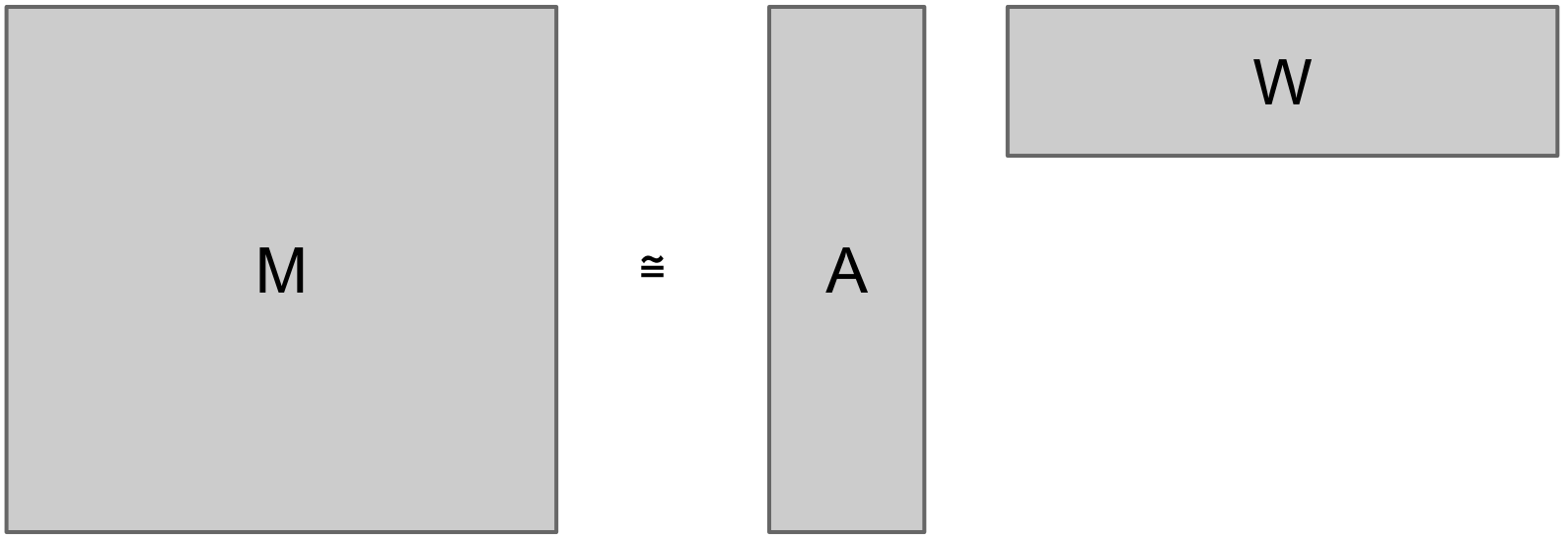
Lecture 2: Matrix Factorizations

Matrix Factorizations

- Basic “structure” for Unsupervised Learning
- Singular Value Decomposition
- Nonnegative Matrix Factorization

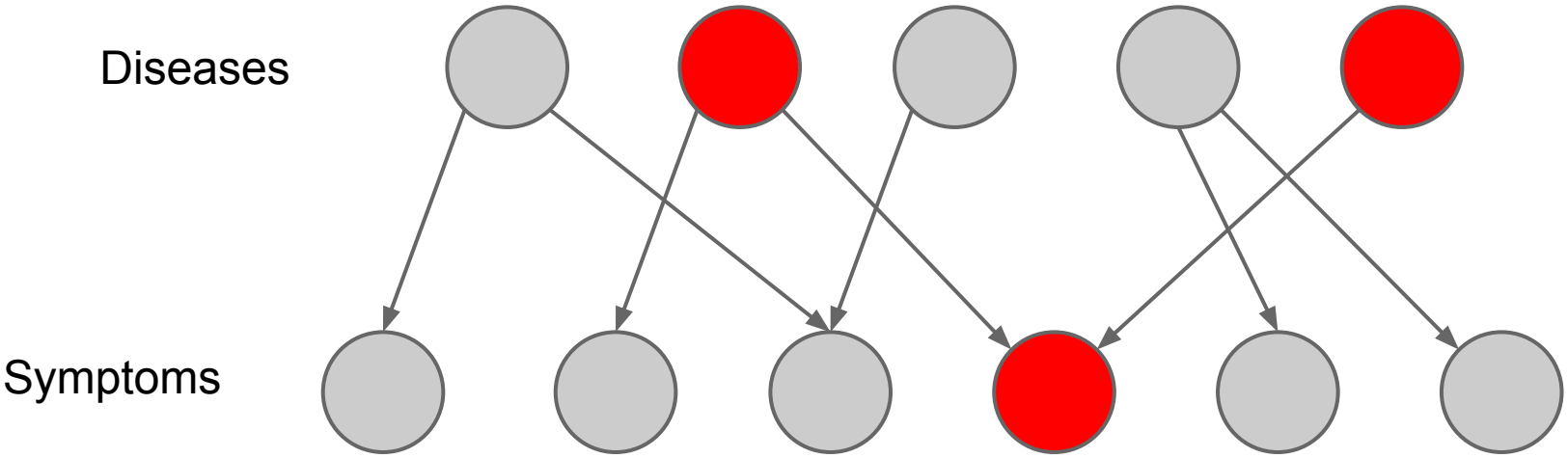
(General) Matrix Factorization



- A, W smaller ⇒ “Simple” for generalization
- Constraints on A, W
  - orthogonal, nonnegative, sparse, ...
- Alternatives for “=”
  - noise, non-linearity, sampling, ...

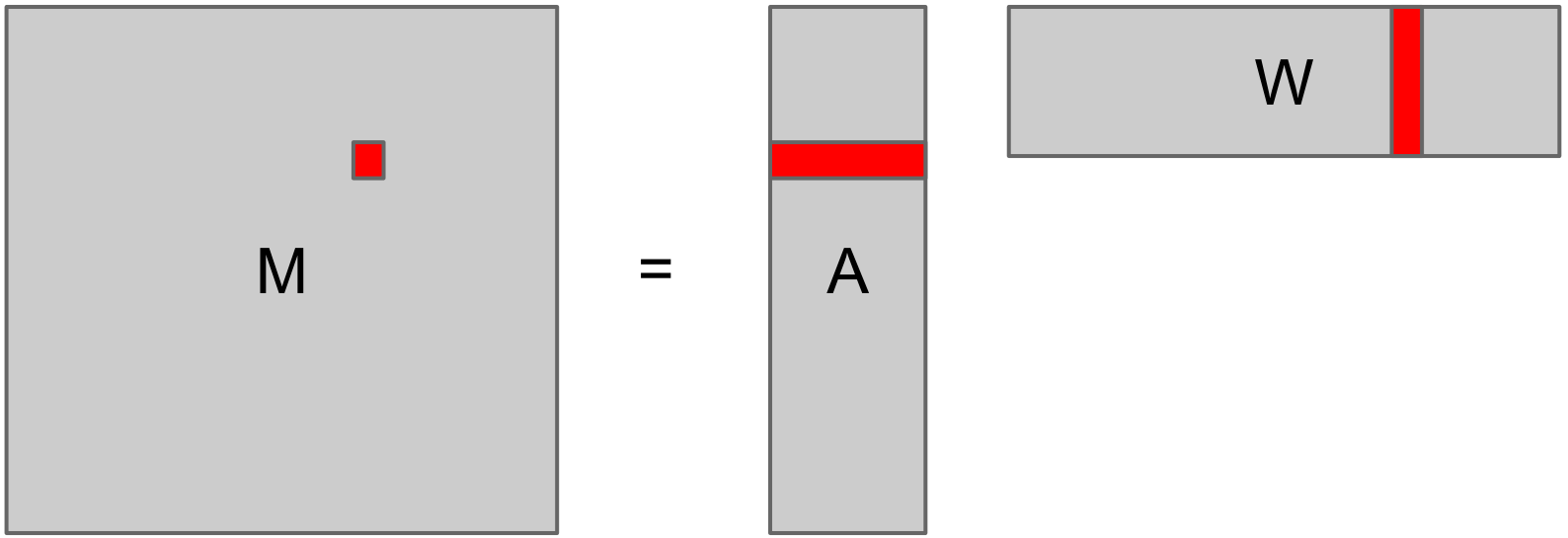
Example: Diseases and Symptoms

QMR-DT network



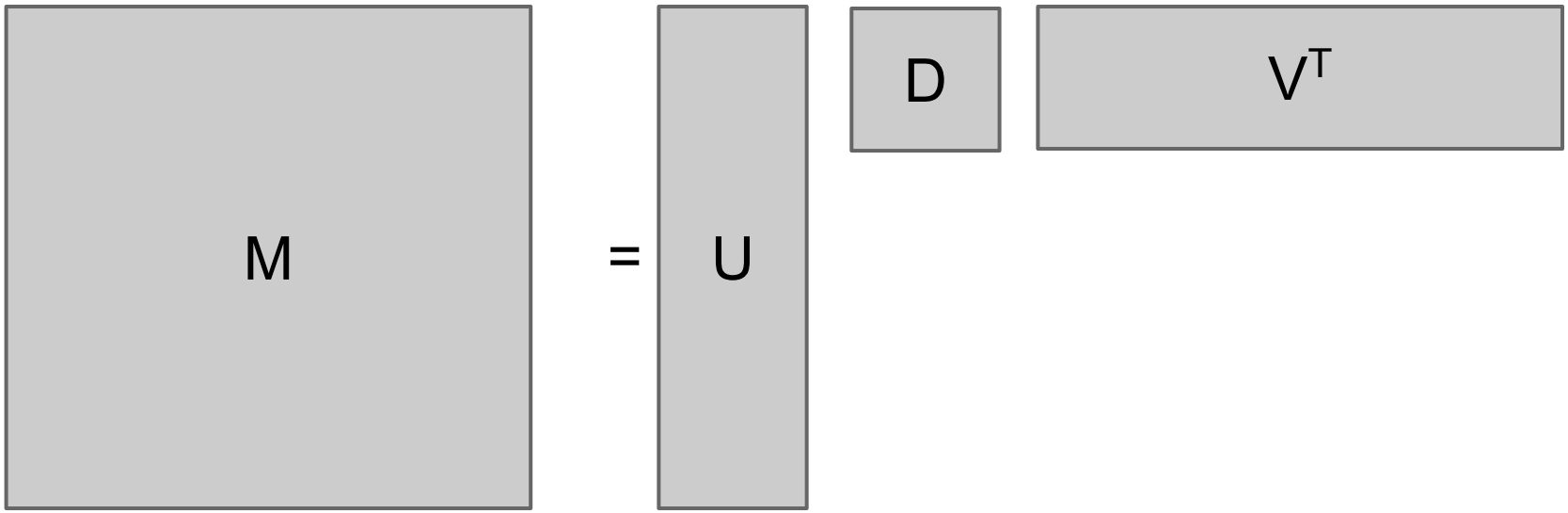
$$\Pr[\text{symptom}] = 1 - \prod (1 - \Pr[\text{symptom}|\text{disease}])$$

Matrix Factorization for QMR



- column of W ⇒ Diseases
- row of A ⇒ log of Pr[no symptom|disease]
- “=” ⇒ sample according to $\exp(-\langle A_i, W_i \rangle)$

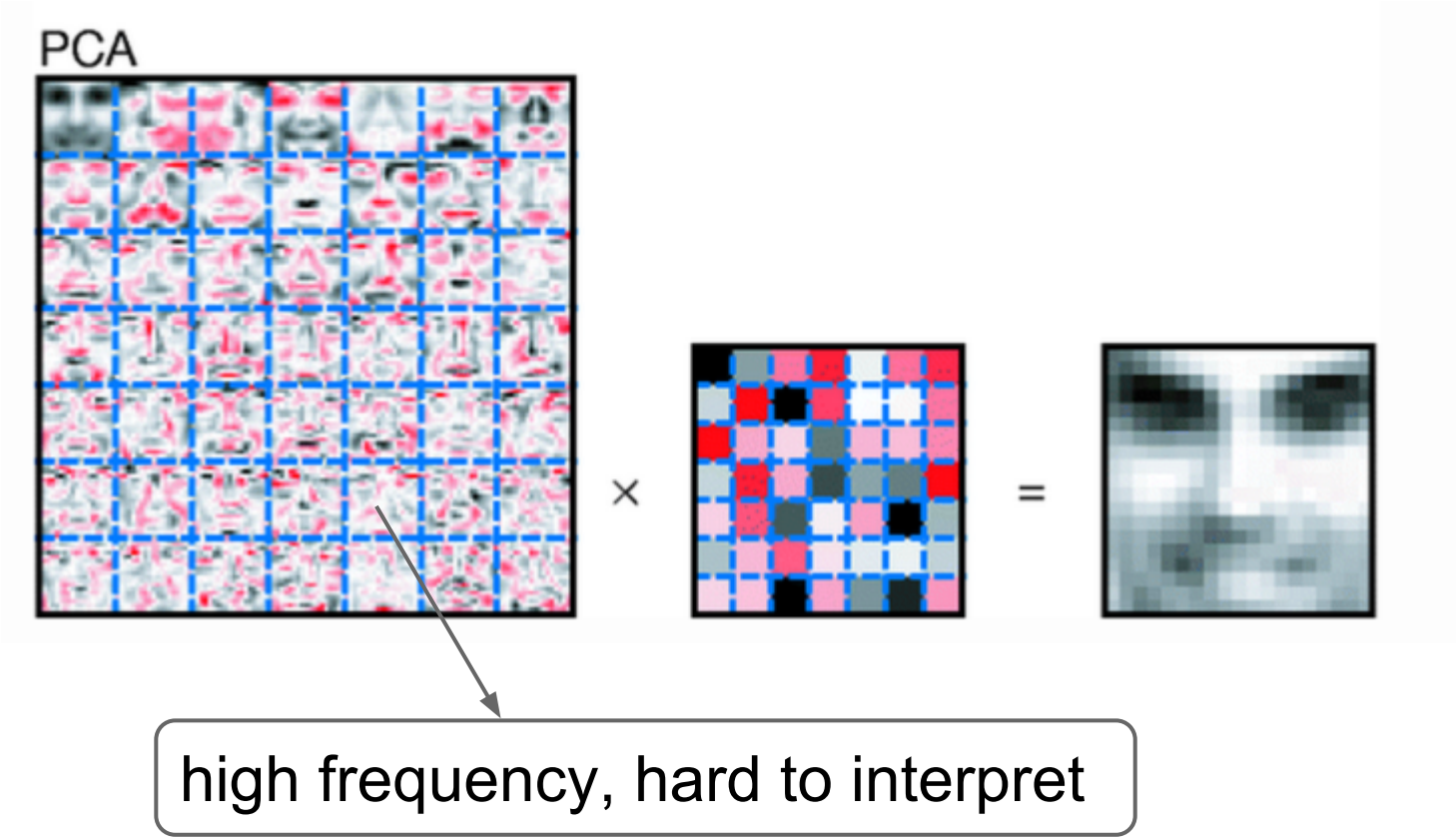
Singular Value Decomposition



- orthogonal components
- efficiently computable
- “optimal” in many ways

Nonnegative Matrix Factorization

[Lee Seung 1999]



Nonnegative Matrix Factorization

[Lee Seung 1999]