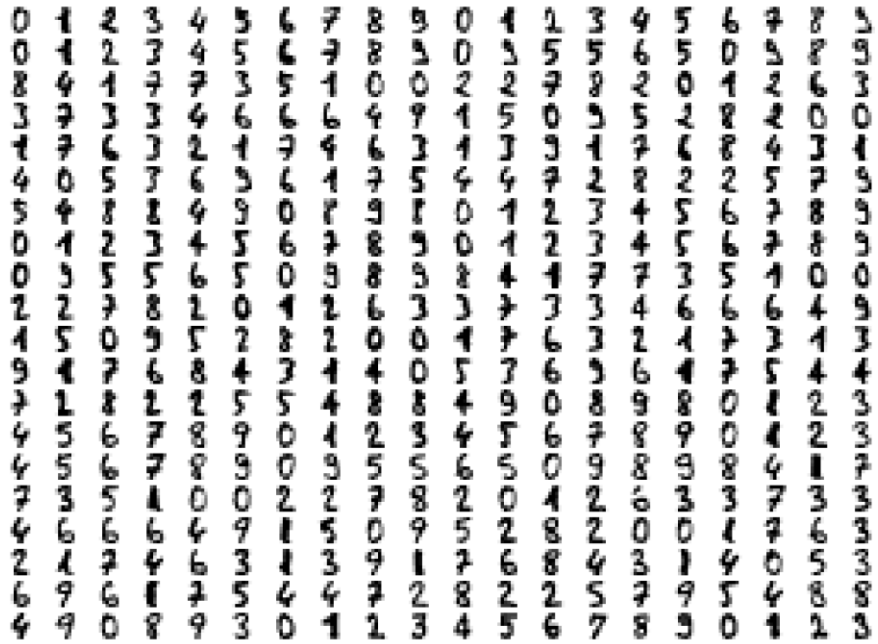
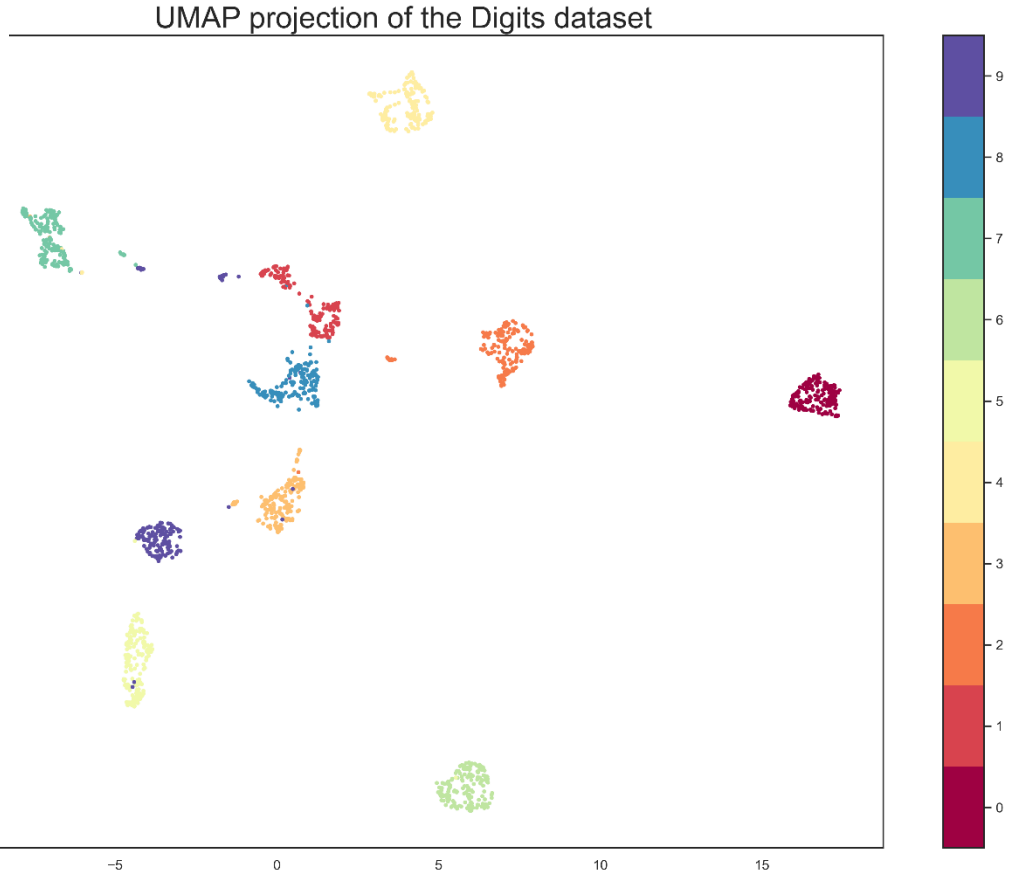
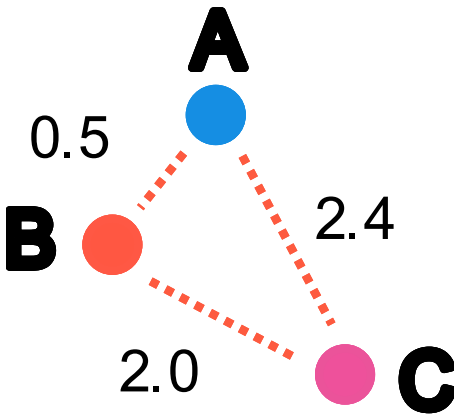
UMAP lecture (Uniform manifold approximation and projection, McInnes 2020)

Alexander Glötzl
Praktische Bioinformatik II
FAKULTÄT FÜR PHYSIK

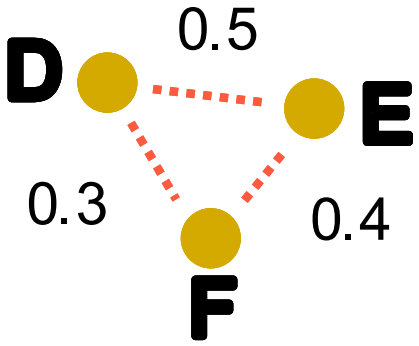
Universität Regensburg

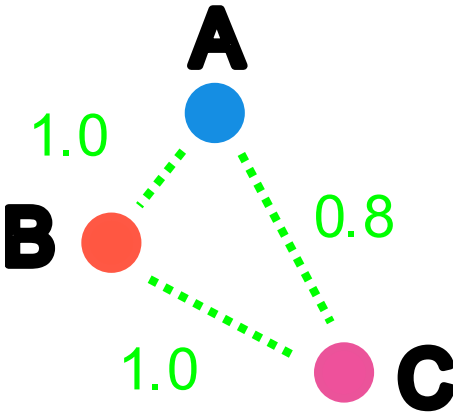



UMAP projection of the Digits dataset
10
5
0
−5
−10
−5
0
5
10
15
9
8
7
6
5
4
3
2
1
0

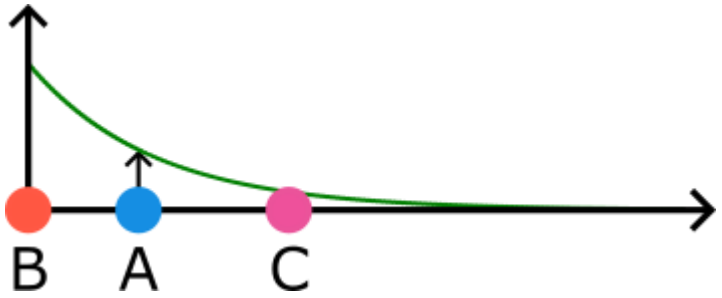


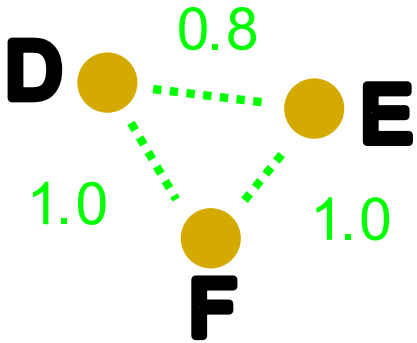
$k = 3$





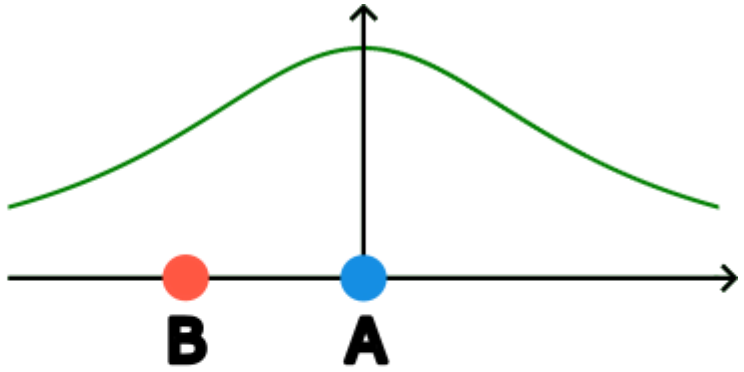
$k = 3$

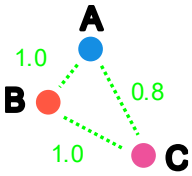


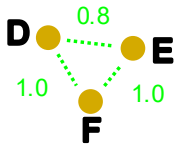


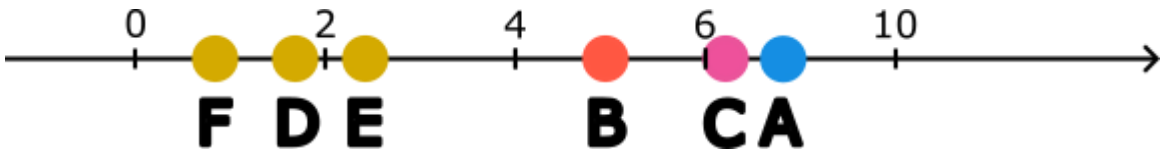
$$\underbrace{\sum_{j=1}^{k} \exp\left(\frac{-(\max(0, d(x_i, x_j) - d_{nearest\ \ neighbor}))}{\sigma_i}\right)}_{w_{high}(x_i, x_j)} = log_2(k)$$

$$w_{low} = \frac{1}{1 + \alpha x^{2\beta}}$$

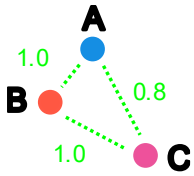


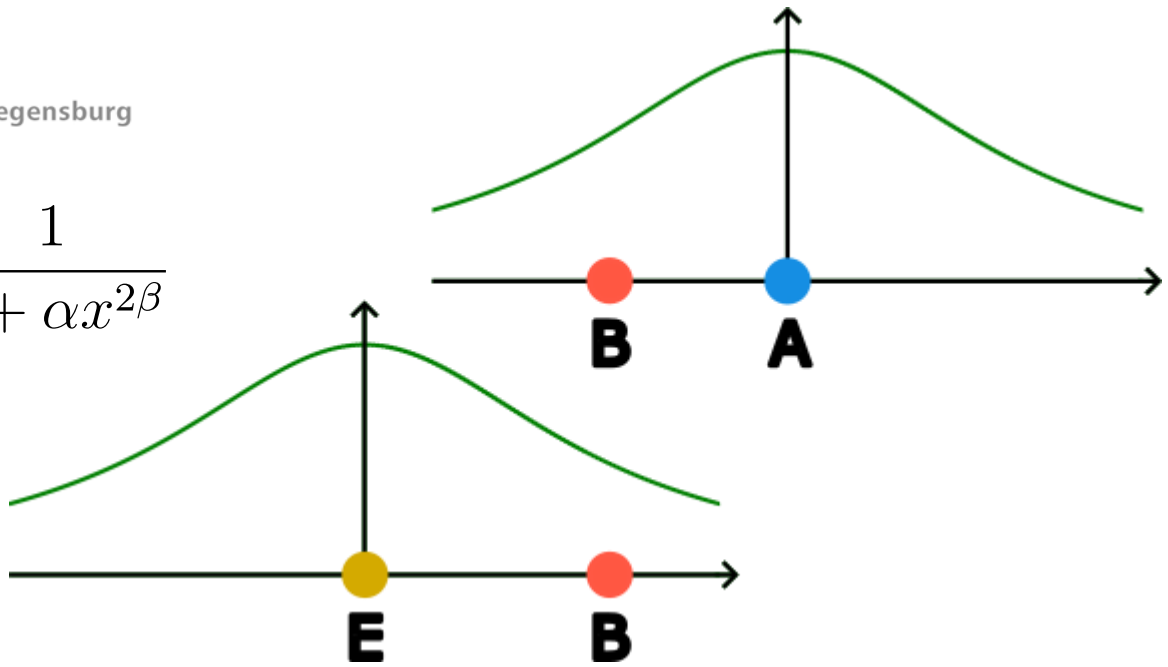


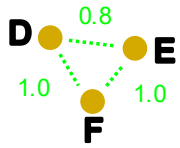


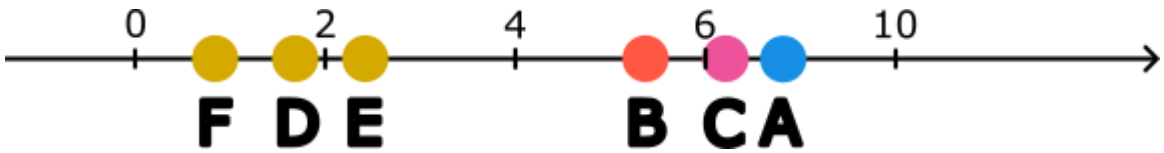


$$w_{low} = \frac{1}{1 + \alpha x^{2\beta}}$$









$$\text{cost} = \sum_{e \in E} \left[w_{high}(e) \cdot log\left(\frac{w_{high}(e)}{w_{low}(e)}\right) + (1 - w_{high}(e)) \cdot log\left(\frac{1 - w_{high}(e)}{1 - w_{low}(e)}\right) \right]$$

Excercises:

- distance matrix (nearest neighbor descent?)
- high dimensional weights/scores via binary search ☑
- spectral embedding ☑
- low dimensional weights/scores ☑
- stochastic gradient descent of entropy