

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

Alexander Glötzl
Praktische Bioinformatik II
FAKULTÄT FÜR PHYSIK

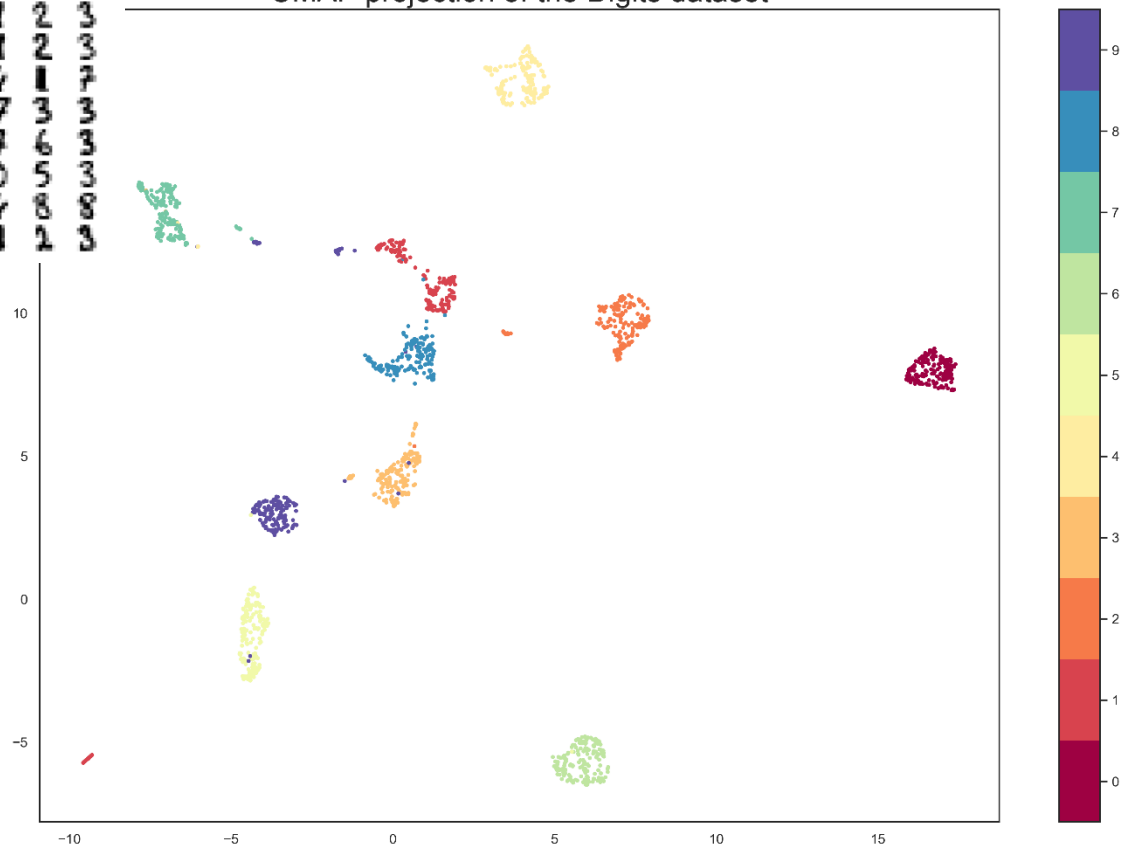


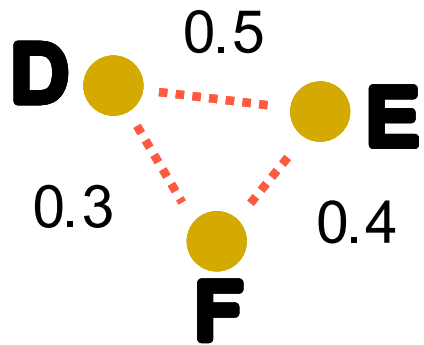
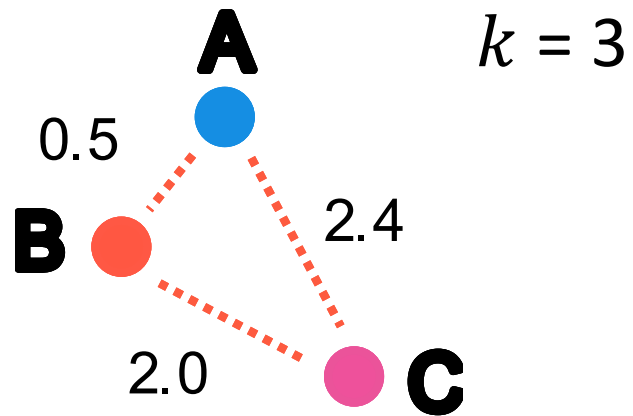
Universität Regensburg

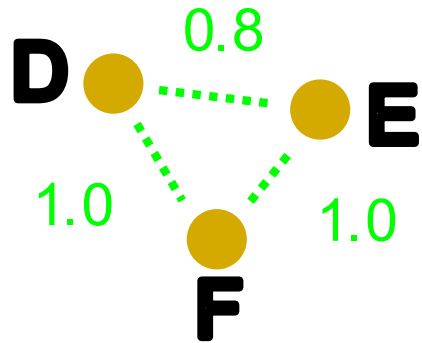
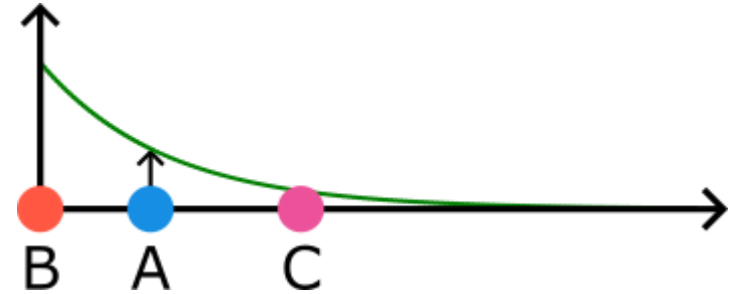
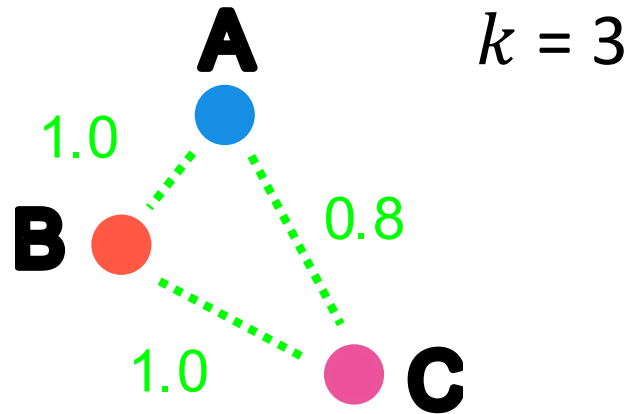
0 0 8 3 1 4 5 0 0 2 1 9 7 4 4 2 6 4
1 1 4 3 1 4 5 0 2 1 9 7 4 4 2 6 4
2 2 1 3 3 0 4 1 3 2 5 4 1 3 5 6 4
3 3 4 3 3 5 8 2 3 5 0 7 8 2 7 4 8
4 4 7 4 2 6 4 4 6 1 5 8 2 8 0 4 7
5 5 3 6 1 3 9 3 5 0 2 4 5 9 9 0 3
6 6 5 6 9 6 0 6 0 4 8 3 5 0 2 1 4 0
7 7 1 6 4 1 7 9 8 6 0 4 4 8 2 5 7 1
8 8 0 4 6 7 9 8 8 6 0 4 4 8 2 5 7 1
9 9 0 9 3 5 1 9 9 3 0 0 8 3 5 8 2 3
0 0 2 1 1 4 4 0 0 8 3 4 5 6 2 8 4
1 1 2 5 3 4 4 1 1 4 7 3 6 0 4 8 2 6
2 2 5 0 9 9 2 2 3 7 3 3 9 8 7 9 5 7
3 3 5 8 9 1 2 3 3 7 3 3 9 8 7 9 5 7
4 4 6 2 5 7 8 4 4 7 4 2 6 9 8 6 0 3 7 8
5 5 0 1 2 6 2 2 5 3 6 4 1 4 8 9 9 3 9 3
6 6 0 1 8 8 2 2 6 6 6 3 7 3 6 3 4 4 5 0
7 7 2 2 4 5 7 7 1 6 3 5 4 4 7 7 0 4 1
8 8 6 0 3 7 8 8 0 4 1 4 2 2 1 3 6 8 2
9 9 3 0 4 5 9 9 0 9 3 4 4 3 3 3 8 3



UMAP projection of the Digits dataset

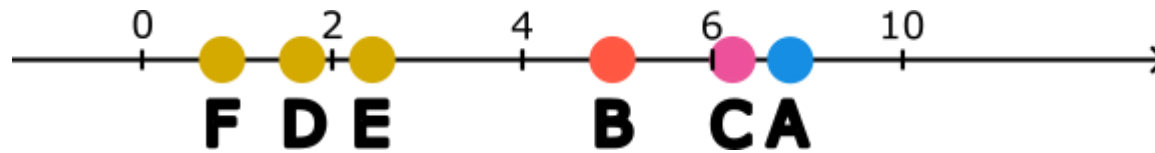
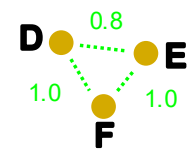
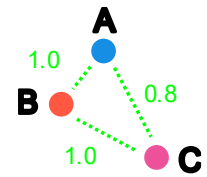
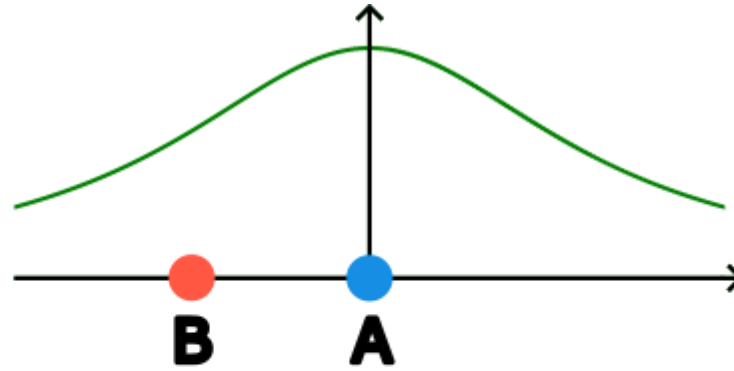




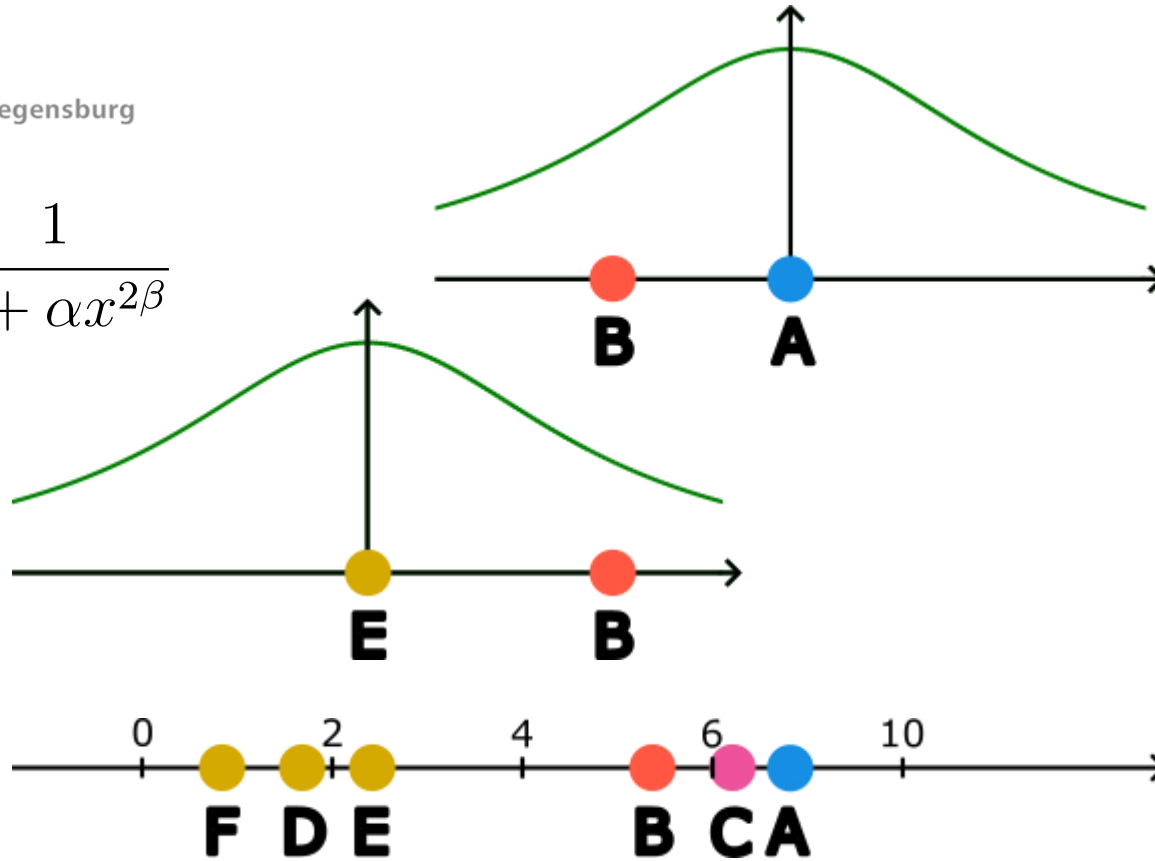
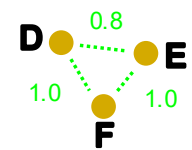
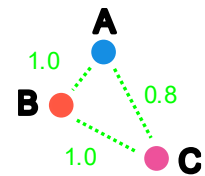


$$\underbrace{\sum_{j=1}^k \exp \left(\frac{-(\max(0, d(x_i, x_j) - d_{\text{nearest neighbor}}))}{\sigma_i} \right)}_{w_{\text{high}}(x_i, x_j)} = \log_2(k)$$

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



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



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

Excercises:

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