

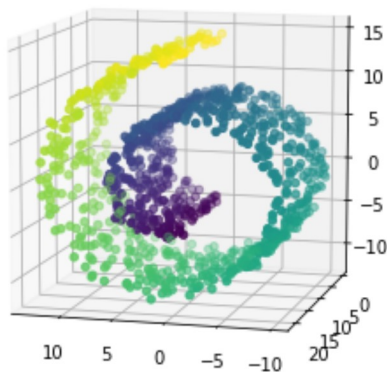
Implementing Locally Linear Embedding (25 P)

In this programming homework we will implement locally linear embedding (LLE) and experiment with it on the swiss roll dataset. In particular, the effects of neighbourhood size and noise on the quality of the embedding will be analyzed.

```
In [1]: import numpy as np
import matplotlib
%matplotlib inline
from matplotlib import pyplot as plt
from mpl_toolkits.mplot3d.axes3d import Axes3D
import sklearn, sklearn.datasets
```

The following code plots the swiss roll dataset (a commonly used dataset to test LLE) with $N=1000$ data points and a noise parameter of 0.25 .

```
In [2]: X, T = sklearn.datasets.make_swiss_roll(n_samples=1000, noise=0.25)
plt.figure(figsize=(5,5))
ax = plt.gca(projection='3d')
ax.view_init(elev=10., azim=105)
ax.scatter(X[:,0], X[:,1], X[:,2], c=T)
plt.show()
```



Although the dataset is in three dimensions, the points follow a two-dimensional low-dimensional structure. The goal of embedding algorithms is to extract this underlying structure, in this case, unrolling the swiss roll into a two-dimensional Euclidean space.

In the following, we consider a simple implementation of LLE. You are required to complete the code by writing the portion where the optimal reconstruction weights are extracted. (*Hint: During computation, you need to solve an equation of the type $Cw=\mathbf{1}$, where $\mathbf{1}$ is a column vector $(1,1,\dots,1)$. In case $k>d$ i.e. the size of the neighbourhood is larger than the number of dimensions of the input space, it is necessary to regularize the matrix C . You can do this by adding positive terms on the diagonal. A good starting point is 0.05 .)*

```

In [64]: import sys
def LLE(X,k):
    N,M = np.shape(X)
    W = np.zeros([N,N])
    #euclidian distance squared, for calculation of nearest neighbours
    dist = lambda x,y: np.sum((x - y) ** 2, axis=0)
    #compute all distances
    D = np.zeros((N,N))
    for i in range(N):
        for j in range(i+1):
            d = dist(X[i,:],X[j,:])
            D[i][j] = d
            D[j][i] = d
    eta = np.zeros((k,M))
    for i in range(N):
        # -----
        # TODO: Replace by your code
        # -----
        #current data point x_i
        x = X[i,:]
        #calculate nearest neighbours, store them in matrix eta, take k+1 becaus
e d(x,x)=0 is also stored in D
        neighb = D[i,:].argsort()[1:k+1]
        for j in range(k):
            eta[j,:] = X[neighb[j],:]
        #calculate C as on homework sheet
        C = np.dot(np.outer(np.ones(k),x) - eta,np.transpose( np.outer(np.ones
(k),x) - eta) )
        #check regularity
        while np.linalg.cond(C) >= 1/sys.float_info.epsilon:
            C = C + np.diag(np.ones(k))*0.05
        Cinv = np.linalg.inv(C)
        #calculate weights, as on homework sheet
        w = Cinv.dot(np.ones(k))/(np.transpose(np.ones(k)).dot(Cinv).dot(np.ones
(k)))
        # -----
        #store weights in matrix W
        for ind in range(w.shape[0]):
            W[i,neighb[ind]] = w[ind]

    M = np.identity(N) - W - W.T + np.dot(W.T,W)
    E = np.linalg.svd(M)[0][:-3:-1]

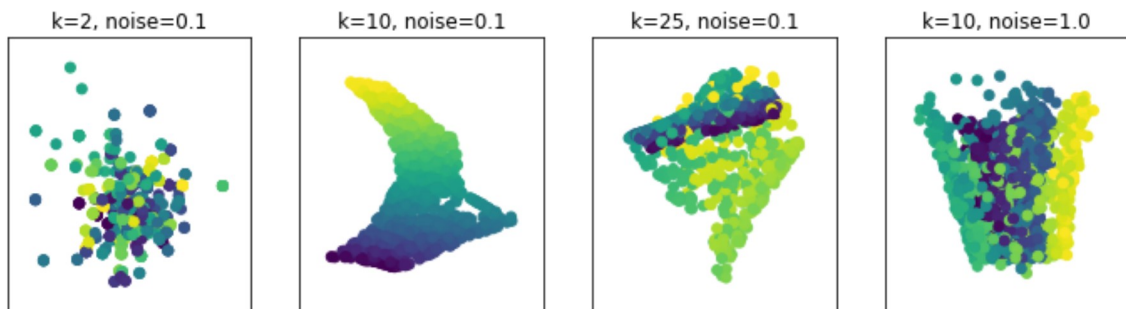
    return E

```

You can now test your implementation on the swiss roll dataset and vary the noise in the data and the parameter k of the LLE algorithm. Results are shown below:

```
In [63]: f = plt.figure(figsize=(12,3))
for t,(k,noise) in enumerate([(2,0.1),(10,0.1),(25,0.1),(10,1)]):
    X,T = sklearn.datasets.make_swiss_roll(n_samples=1000, noise=noise)
    embedding = LLE(X,k=k)
    ax = f.add_subplot(1,4,t+1)
    ax.set_title('k=%d, noise=%.1f'%(k,noise))
    ax.set_xticks([])
    ax.set_yticks([])
    ax.scatter(embedding[:,0],embedding[:,1],c=T)
f.show()
```

C:\Users\UltraBook\anaconda3\lib\site-packages\ipykernel_launcher.py:10: UserWarning: Matplotlib is currently using module://ipykernel.pylab.backend_inline, which is a non-GUI backend, so cannot show the figure.
 # Remove the CWD from sys.path while we load stuff.



It can be observed that the parameter k must be carefully tuned to have sufficiently many neighbors for stability but also not too many. We can further observe that LLE works well as long as the noise in the data remains low enough.