Homework 8

1.

1.1 设数据为 X. 减去均值后为 $\tilde{X} \in \mathbb{R}^{D\times N}$ covariance 为 $S = \tilde{X}\tilde{X}^T$

投影到一维方向 u_1 得到的数据矩阵为 $u_1^T \tilde{X}$

variance 为 $u_1^T \tilde{X} \tilde{X}^T u_1 = u_1^T S u_1$

1.2 引入限制 $u_1^T u_1 = 1$ $\max_{\|u_1\|_2=1} u_1^T S u_1$

introduce Lagranian multipler $\lambda \Rightarrow L = u_1^T S u_1 + \lambda(u_1^T u_1 - 1)$

$\frac{\partial L}{\partial u_1} = 2Su_1 + 2\lambda u_1 = 0 \Rightarrow Su_1 = -\lambda u_1$

u_1 一定为 S 的特征值（对应 $-\lambda$），代回 $u_1^T S u_1 = (-\lambda) u_1^T u_1 = (-\lambda)$

因需让 $(-\lambda)$ 最大 即 u_1 为对应 S 最大特征值的特征向量

1.3 设已求得前 M 个最好方向均为 S 的前 M 个最大特征值. 求第 $M+1$ 方向.

$$L = u_{M+1}^T S u_{M+1} + \lambda(u_{M+1}^T u_{M+1} - 1) + \sum_{i=1}^{M} v_i u_i^T u_{M+1}$$

$$\frac{\partial L}{\partial u_{M+1}} = 2S u_{M+1} + 2\lambda u_{M+1} + \sum_{i=1}^{M} v_i u_i$$

左乘 u_{M+1}^T $u_{M+1}^T S u_{M+1} + \lambda u_{M+1}^T u_{M+1} = 0$

$$\lambda = \frac{u_{M+1}^T S u_{M+1}}{u_{M+1}^T u_{M+1}} = u_{M+1}^T S u_{M+1}$$

左乘 u_j^T 有 $2u_j^T S u_{M+1} + v_i = 0$

$\Rightarrow 2\lambda_j u_j^T u_{M+1} + v = 0$

$\Rightarrow v = 0$

$\therefore \frac{\partial L}{\partial u_{M+1}} = 0 \Rightarrow S u_{M+1} = (-\lambda) u_{M+1}$

$\therefore$ u_{M+1} 为 S 特征值 且与 $u_1 \cdots u_M$ 垂直

且要最大化 $u_{M+1}^T S u_{M+1}$

$\therefore$ u_{M+1} 为 S 的第 $M+1$ 个特征向量

1.4 $J = \frac{1}{N}\sum_{n=1}^{N} \| x_n - \sum_{i=1}^{M} z_{ni} u_i - \sum_{i=M+1}^{D} b_i u_i \|^2$

$$\frac{\partial J}{\partial z_{ni}} = \frac{1}{N} 2 \left(x_n - \sum_{j=1}^{M} z_{nj} u_j - \sum_{j=M+1}^{D} b_j u_j\right)^T u_i = 0$$

$u_j^T u_i = 0 \ (j \neq i)$

$$\Rightarrow x_n^T u_i - z_{ni} u_i^T u_i = 0 \qquad (i = 1, \cdots M)$$

$u_i^T u_i = 1$

$$\Rightarrow z_{ni} = x_n^T u_i$$

$$\frac{\partial J}{\partial b_i} = \frac{1}{N} \sum_{n=1}^{N} 2 \left(x_n - \sum_{j=1}^{M} z_{nj} u_j - \sum_{j=M+1}^{D} b_j u_j\right)^T u_i = 0$$

$$\Rightarrow \sum_{n=1}^{N} x_n^T u_i - \sum_{n=1}^{N} b_i u_i^T u_i = 0$$

$$\Rightarrow b_i = \left(\frac{\sum_{n=1}^{N} x_n}{N}\right)^T u_i = \bar{x}^T u_i$$

1.5 $x_n - \hat{x}_n = \sum_{i=M+1}^{D} \{(x_n - \bar{x})^T u_i\} u_i$

$$J = \frac{1}{N}\sum_{n=1}^{N} \|x_n - \hat{x}_n\|^2 = \frac{1}{N} \sum_{n=1}^{N} \left(\sum_{i=M+1}^{D} \{(x_n - \bar{x})^T u_i\} u_i\right)^T \left(\sum_{i=M+1}^{D} \{(x_n - \bar{x})^T u_i\} u_i\right)$$

$$\overset{u_i u_j = 0 \ (i \neq j)}{=} \frac{1}{N} \sum_{n=1}^{N} \sum_{i=M+1}^{D} u_i^T (x_n - \bar{x})(x_n - \bar{x})^T u_i \, u_i^T u_i$$

$$\overset{u_i^T u_i = 1}{=} \frac{1}{N} \sum_{n=1}^{N} \sum_{i=M+1}^{D} u_i^T (x_n - \bar{x})(x_n - \bar{x})^T u_i$$

$$= \sum_{i=M+1}^{D} u_i^T S u_i = \mathrm{Tr}\{U^T S U\} \qquad U \in \mathbb{R}^{D \times (D-M)}$$

constraint: $u_i u_j = \delta_{ij}$ 即 $U^T U = I_{D-M}$ 并对每维加一个 langranian multiplier

$$\therefore L = \mathrm{Tr}\{U^T S U\} + \lambda^T \mathrm{vec}(U^T U - I)$$

vec为将矩阵转为一维向量形式

$$= \mathrm{Tr}\{U^T S U\} + \mathrm{Tr}\{H(\overset{I - U^T U}{\cancel{U^T U - I}})\}$$

其中 $\lambda = \mathrm{vec}(H)$

1.6 $\frac{\partial L}{\partial U} = SU + S^TU - (UH^T + UH) = 0$

H为对称阵，因为其对应约束为同一约束

∴有 $SU = UH \Rightarrow H = U^TSU$

重建误差 $Tr\{U^TSU\} = Tr\{H\}$

∴ u_i 为S特征向量，$h_{i,i}$ 为对应特征值

$\sum_{i=M+1}^{D} h_{i,i} = Tr\{H\}$ 为最小化目标

$SU_{M+1} = \sum_{i=M+1}^{D} h_{M+1,i} u_i$

同乘 u_{M+1}^T u_{M+1}

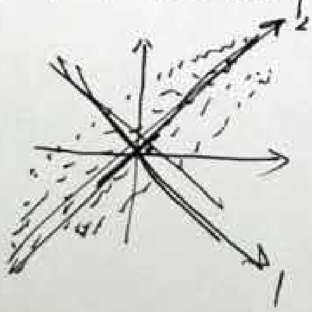
$\Rightarrow SU_{M+1} = h_{M+1,M+1} u_{M+1}$

同理可得 $Su_i = h_{i,i} u_i \quad (i=M+1, \cdots D)$

∴应取 u_i $(i=M+1, \cdots D)$ 为 S 的最小的 D-M 个特征向量.

此时 $Tr\{H\} = \sum_{i=M+1}^{D} h_{i,i}$ 为S的最小的 D-M 个特征值之和.

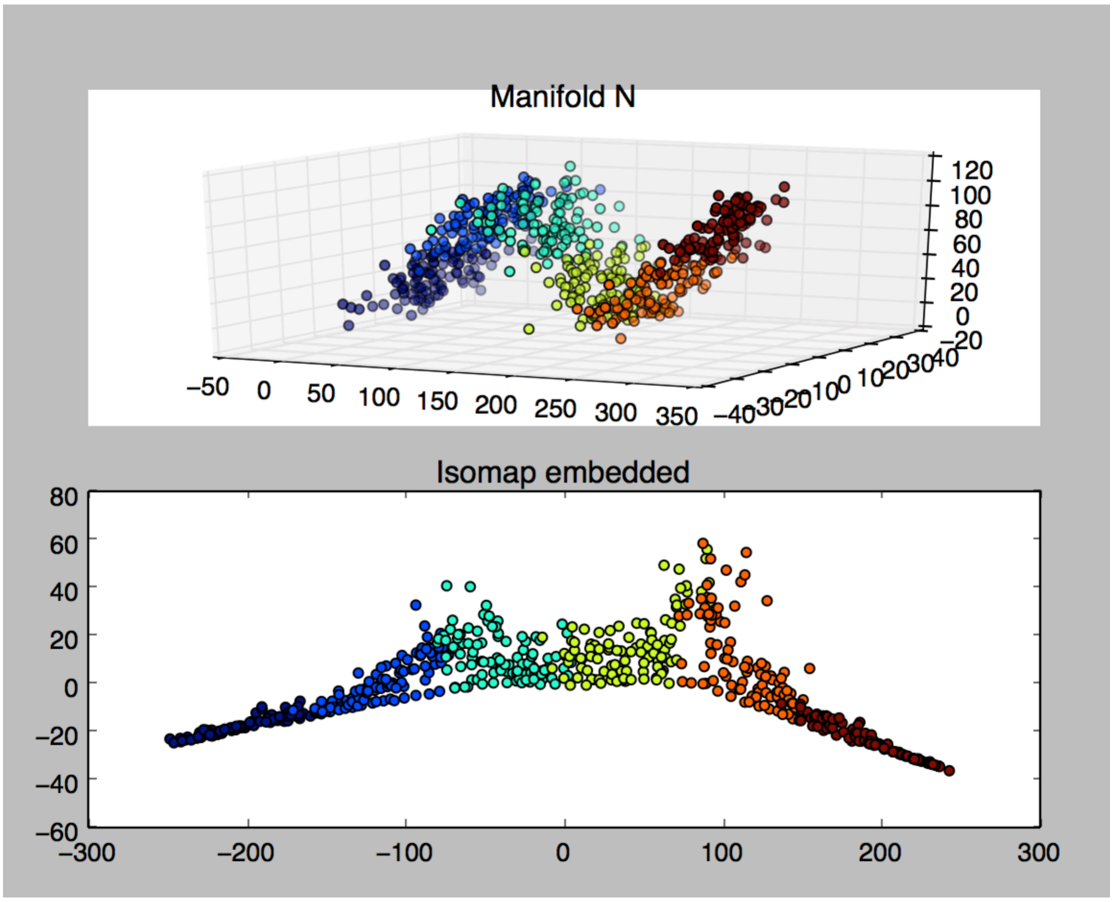
1.7 假设二维数据分布(已centered)在2方向variance较大，1方向variance较小



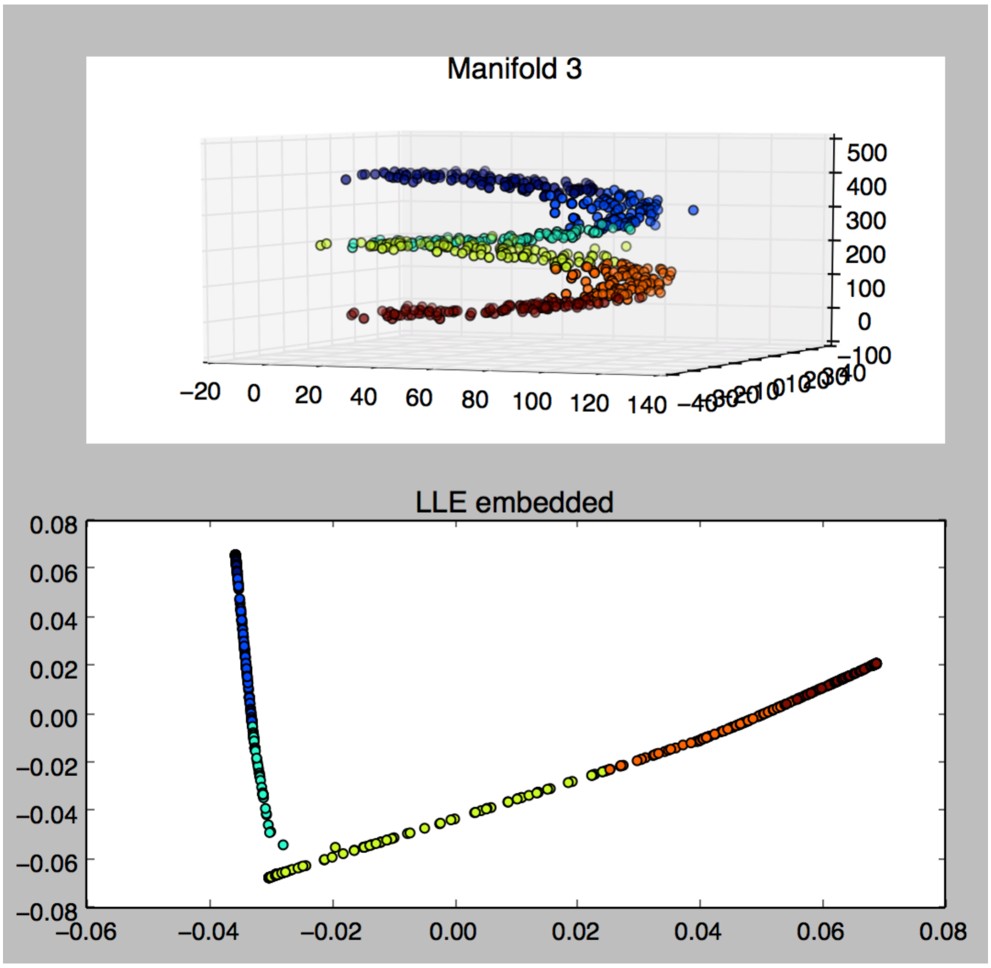
若选一个方向使投影数据variance最大、应选2方向

若希望投影后~~向量~~向量与原向量误差最小，也应选2方向.

Isomap 对 N 流形的结果，取 neighbor 数为 5. 得到结果如下：



LLE 对 3 流形的结果，由于本来高维空间也只有 3 维，取邻居为 5 个邻居，在高维空间就基本都可以用 5 个邻居完全 fit 所有节点，这样在低维空间得到的 fit 结果会很差，所以需要适当加大 reguliarized term，这里设置为 0.1.


Manifold 3
LLE embedded