

## LLE - Locally Linear Embedding

---

Advanced Institute for Artificial Intelligence – AI2

<https://advancedinstitute.ai>

A técnica *Locally Linear Embedding* (LLE) é uma técnica **local** para redução não linear de dimensionalidade que baseia-se no arcabouço de **aprendizado de variedades** (*manifold learning*). A principal diferença desse método para ISOMAP é que este é uma técnica **global**, pois usávamos a matriz de distâncias entre **todas as amostras** do conjunto de dados para aprendermos o novo espaço reduzido. No caso do LLE, iremos considerar apenas a vizinhança (*patch*) de um dado ponto para aprendermos a sua nova coordenada no espaço reduzido.

Hipótese: caso tomemos uma amostra e os seus vizinhos mais próximos, teremos um *patch* linear (espaço Euclidiano), ou seja, eu posso aproximar o elemento central (amostra) por meio da **combinação linear** dos seus vizinhos.

Seja  $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m\}$  o nosso conjunto de dados original tal que  $\mathbf{x}_i \in \mathbb{R}^n$ . Temos, então que a aproximação  $\tilde{\mathbf{x}}_i$  da amostra  $\mathbf{x}_i$  pode ser obtida da seguinte forma:

$$\tilde{\mathbf{x}}_i \approx \sum_{\mathbf{x}_j \in \mathcal{N}(\mathbf{x}_i)} w_{ij} \mathbf{x}_j, \quad (1)$$

em que  $\mathcal{N}(\mathbf{x}_i)$  denota a vizinhança (*patch*) da amostra  $\mathbf{x}_i$  e  $w_{ij}$  o peso atribuído durante a combinação linear. O algoritmo do LLE requer, basicamente, o conjunto de dados representado por uma matriz  $\mathbf{X} \in \mathbb{R}^{m \times n}$ , o número desejado de dimensões  $d < n$  e um valor  $k > d + 1$  para o tamanho da vizinhança. Sem perda de generalidade, vamos assumir que  $\mathbf{X}_i = \mathbf{x}_i$ , ou seja, a  $i$ -ésima linha da matriz  $\mathbf{X}$  corresponde à  $i$ -ésima amostra de  $\mathcal{X}$ .

O algoritmo do LLE pode ser dividido em três passos principais:

- ❶ Para cada  $\mathbf{x}_i \in \mathcal{X}$  encontrar os seus  $k$ -vizinhos mais próximos,  $i = 1, 2, \dots, m$ .
- ❷ Encontrar a matriz  $\mathbf{W}$  que minimiza o erro de reconstrução para cada amostra  $\mathbf{x}_i \in \mathcal{X}$ :

$$E(\mathbf{W}, \mathbf{X}) = \sum_{i=1}^m \left\| \mathbf{x}_i - \sum_{\mathbf{x}_j \in \mathcal{N}(\mathbf{x}_i)} w_{ij} \mathbf{x}_j \right\|^2, \quad (2)$$

em que  $w_{ij} = 0$  a menos que  $\mathbf{x}_j \in \mathcal{N}(\mathbf{x}_i)$ . Note que  $\sum_{\mathbf{x}_j \in \mathcal{N}(\mathbf{x}_i)} w_{ij} = 1$ .

- ❸ Encontre o conjunto de coordenadas  $\hat{\mathbf{X}}$  que minimiza o erro de reconstrução utilizando os pesos ótimos encontrados no passo anterior:

$$\Phi(\mathbf{W}, \hat{\mathbf{X}}) = \sum_{i=1}^m \left\| \hat{\mathbf{x}}_i - \sum_{\mathbf{x}_j \in \mathcal{N}(\mathbf{x}_i)} w_{ij} \hat{\mathbf{x}}_j \right\|^2, \quad (3)$$

tal que  $\hat{\mathbf{x}} \in \mathbb{R}^d$  e  $\hat{\mathbf{X}} \hat{\mathbf{X}}^T = \mathbf{I}$ . Sem perda de generalidade, vamos assumir que  $\hat{\mathbf{X}}_i = \hat{\mathbf{x}}_i$ , ou seja, a  $i$ -ésima linha da matriz  $\hat{\mathbf{X}}$  corresponde à  $i$ -ésima amostra de  $\hat{\mathcal{X}}$ .

O primeiro passo corresponde à encontrar os  $k$ -vizinhos mais próximos de  $\mathbf{x}_i$ ,  $\forall \mathbf{x}_i \in \mathcal{X}$ . Em geral, valores pequenos de  $k$  tornam o grafo desconexo, enquanto que um  $k > n$  exige uma **regularização** do problema.

Agora, vamos para o passo 2, ou seja, estimação dos pesos  $\mathbf{W}$  utilizando os mínimos quadrados. Sem perda de generalidade, podemos expressar o erro de reconstrução local considerando apenas o ponto  $\mathbf{x}_i \in \mathcal{X}$  como segue:

$$\begin{aligned}
 E(\mathbf{w}_i, \mathbf{x}_i) &= \left\| \mathbf{x}_i - \sum_{\mathbf{x}_j \in \mathcal{N}(\mathbf{x}_i)} w_{ij} \mathbf{x}_j \right\|^2 = \left\| \sum_{\mathbf{x}_j \in \mathcal{N}(\mathbf{x}_i)} w_{ij} \mathbf{x}_i - \sum_{\mathbf{x}_j \in \mathcal{N}(\mathbf{x}_i)} w_{ij} \mathbf{x}_j \right\|^2 \\
 &= \left\| \sum_{\mathbf{x}_j \in \mathcal{N}(\mathbf{x}_i)} w_{ij} (\mathbf{x}_i - \mathbf{x}_j) \right\|^2 = \sum_{\mathbf{x}_j \in \mathcal{N}(\mathbf{x}_i)} \sum_{\mathbf{x}_k \in \mathcal{N}(\mathbf{x}_i)} w_{ij} w_{ik} (\mathbf{x}_i - \mathbf{x}_j)^T (\mathbf{x}_i - \mathbf{x}_k). \quad (4)
 \end{aligned}$$

*Note: In the original image, a red arrow points from the text "1 (restrição)" to the term  $w_{ij} \mathbf{x}_i$  in the second norm expression.*

Vamos, agora, definir a matriz  $\mathbf{C} \in \mathbb{R}^{k \times k}$ :

$$C_{jk} = (\mathbf{x}_i - \mathbf{x}_j)^T (\mathbf{x}_i - \mathbf{x}_k). \quad (5)$$

Desta forma, podemos reescrever a expressão para o erro local de reconstrução (Equação 4) como segue:

$$E(\mathbf{w}_i, \mathbf{x}_i) = \sum_j \sum_k w_{ij} C_{jk} w_{ik} = \mathbf{w}^T \mathbf{C} \mathbf{w}. \quad (6)$$

Lembrando que o intuito principal da restrição  $\sum_{\mathbf{x}_j \in \mathcal{N}(\mathbf{x}_i)} w_{ij} = 1$  é para adicionar propriedades de **invariância à translação**, isto é, adicionando uma constante  $c \in \mathbb{R}^n$  ao vetor  $\mathbf{x}_i$  e aos seus vizinhos, o erro de reconstrução não muda.

Sejam, então,  $\mathbf{x}'_i = \mathbf{x}_i + \mathbf{c}$  e  $\mathbf{x}'_j = \mathbf{x}_j + \mathbf{c}$  as versões deslocadas das amostras  $\mathbf{x}_i$  e  $\mathbf{x}_j$ , respectivamente. O novo erro de reconstrução local é dado por:

$$E'(\mathbf{w}_i, \mathbf{x}_i) = \left\| \mathbf{x}'_i - \sum_{\mathbf{x}_j \in \mathcal{N}(\mathbf{x}_i)} w_{ij} \mathbf{x}'_j \right\|^2 = \left\| \mathbf{x}_i + \mathbf{c} - \sum_{\mathbf{x}_j \in \mathcal{N}(\mathbf{x}_i)} w_{ij} (\mathbf{x}_j + \mathbf{c}) \right\|^2. \quad (7)$$

Podemos reescrever esta equação como segue:

$$\begin{aligned}
E'(\mathbf{w}_i, \mathbf{x}_i) &= \left\| \mathbf{x}'_i - \sum_{\mathbf{x}_j \in \mathcal{N}(\mathbf{x}_i)} w_{ij} \mathbf{x}'_j \right\|^2 = \left\| \mathbf{x}_i + \mathbf{c} - \sum_{\mathbf{x}_j \in \mathcal{N}(\mathbf{x}_i)} w_{ij} (\mathbf{x}_j + \mathbf{c}) \right\|^2 \\
&= \left\| \mathbf{x}_i + \mathbf{c} - \sum_{\mathbf{x}_j \in \mathcal{N}(\mathbf{x}_i)} w_{ij} \mathbf{x}_j - \sum_{\mathbf{x}_j \in \mathcal{N}(\mathbf{x}_i)} w_{ij} \mathbf{c} \right\|^2 \\
&= \left\| \mathbf{x}_i + \mathbf{c} - \sum_{\mathbf{x}_j \in \mathcal{N}(\mathbf{x}_i)} w_{ij} \mathbf{x}_j - \mathbf{c} \sum_{\mathbf{x}_j \in \mathcal{N}(\mathbf{x}_i)} w_{ij} \right\|^2 \\
&= \left\| \mathbf{x}_i + \mathbf{c} - \sum_{\mathbf{x}_j \in \mathcal{N}(\mathbf{x}_i)} w_{ij} \mathbf{x}_j - \mathbf{c} \right\|^2 = \left\| \mathbf{x}_i - \sum_{\mathbf{x}_j \in \mathcal{N}(\mathbf{x}_i)} w_{ij} \mathbf{x}_j \right\|^2 = E(\mathbf{w}_i, \mathbf{x}_i).
\end{aligned} \tag{8}$$

Assim, a translação não afeta o erro quando temos a restrição  $\sum_{\mathbf{x}_j \in \mathcal{N}(\mathbf{x}_i)} w_{ij} = 1$ .



Note que a Equação 6 vale para apenas uma variável  $\mathbf{x}_i \in \mathcal{X}$ . Precisamos, então, generalizá-la para todo o conjunto de dados, pois assim teremos  $m$  problemas de otimização com restrição de igualdade, ou seja:

$$\begin{aligned} \mathbf{w}_i^* &= \arg \min_{\mathbf{w}_i} \mathbf{w}_i^T \mathbf{C}_i \mathbf{w}_i, \\ \text{s.a. } \mathbf{1}^T \mathbf{w}_i &= 1, \quad i = 1, 2, \dots, m. \end{aligned} \tag{9}$$

Como temos um problema de otimização com restrições, precisamos fazer uso da técnica dos multiplicadores de Lagrange. Desta forma, a Equação 8 pode ser reformulada da seguinte maneira:

$$L(\mathbf{w}_i, \lambda) = \mathbf{w}_i^T \mathbf{C}_i \mathbf{w}_i - \lambda (\mathbf{1}^T \mathbf{w}_i - 1). \tag{10}$$

Para resolvermos o problema de otimização, basta calcular a derivada da Equação 9 em relação à  $\mathbf{w}_i$  e igualar à 0, ou seja:

$$\frac{\partial L(\mathbf{w}_i, \lambda)}{\partial \mathbf{w}_i} = 2\mathbf{C}_i\mathbf{w}_i - \lambda\mathbf{1}^T = 0. \quad (11)$$

Da equação acima, temos que:

$$\mathbf{C}_i\mathbf{w}_i = \frac{\lambda}{2}\mathbf{1}. \quad (12)$$

Caso a matriz  $\mathbf{C}_i$  seja admita inversa, temos uma equação fechada para nosso problema de otimização:

$$\mathbf{w}_i = \frac{\lambda}{2}\mathbf{C}_i^{-1}\mathbf{1}. \quad (13)$$

Usualmente, descartamos o termo  $\frac{\lambda}{2}$  da Equação 12, pois é uma constante (não depende de  $i$ ). Assim sendo, podemos reescrever a Equação 12 da seguinte forma:

$$C_i \mathbf{w}_i = \mathbf{1}, \quad (14)$$

que é a formulação de um sistema linear. Desta forma, basta resolvermos este sistema utilizando alguma técnica qualquer e, após encontrar  $\mathbf{w}_i$ , realizamos a sua normalização, ou seja:

$$w_{ij} = \frac{w_{ij}}{\sum_j w_{ij}}. \quad (15)$$

Desta forma, garantimos que a restrição  $\sum_j w_{ij} = 1$  será satisfeita.

Feito isso, resolvemos o subproblema 1 para o **caso padrão**, ou seja, quando  $k < n$ . Quando temos a situação oposta, ou seja,  $k > n$ , precisamos **regularizar** o nosso problema, pois o mesmo fica **mal condicionado**, ou seja, existem mais variáveis desconhecidas ( $k$ ) do que o número de equações para resolvê-las ( $n$ ).

Uma técnica bastante comum é a regularização de Tikonov, ou seja, ao invés de minimizar diretamente a equação abaixo:

$$E(\mathbf{w}_i, \mathbf{x}_i) = \left\| \mathbf{x}_i - \sum_{\mathbf{x}_j \in \mathcal{N}(\mathbf{x}_i)} w_{ij} \mathbf{x}_j \right\|^2, \quad (16)$$

adicionamos um termo de regularização:

$$E(\mathbf{w}_i, \mathbf{x}_i) = \left\| \mathbf{x}_i - \sum_{\mathbf{x}_j \in \mathcal{N}(\mathbf{x}_i)} w_{ij} \mathbf{x}_j \right\|^2 + \alpha \sum_{\mathbf{x}_j \in \mathcal{N}(\mathbf{x}_i)} w_{ij}^2. \quad (17)$$

Na equação anterior, temos que  $\alpha$  controla o grau de regularização. A ideia é encontrar uma relação custo-benefício para o valor de  $\alpha$ , ou seja, quando  $\alpha \rightarrow 0$ , temos que o problema torna-se o de mínimos quadrados. Quando  $\alpha \rightarrow \infty$ , o termo de regularização "pesa" bastante no processo de otimização, ou seja, queremos minimizar a norma Euclidiana do vetor  $\mathbf{w}$  de tal forma a anular o termo de regularização. Tipicamente, utilizamos valores de  $\alpha > 0$ , porém pequenos.

Neste caso, o problema de otimização passa a ser o seguinte:

$$\begin{aligned} \mathbf{w}_i^* = \arg \min_{\mathbf{w}_i} & \mathbf{w}_i^T \mathbf{C}_i \mathbf{w}_i + \alpha \mathbf{w}_i^T \mathbf{w}_i, \\ \text{s.a. } & \mathbf{1}^T \mathbf{w}_i = 1, \quad i = 1, 2, \dots, m. \end{aligned} \tag{18}$$

A função Lagrangiana referente à Equação 17 é dada por:

$$L(\mathbf{w}_i, \lambda) = \mathbf{w}_i^T \mathbf{C}_i \mathbf{w}_i + \alpha \mathbf{w}_i^T \mathbf{w}_i - \lambda (\mathbf{1}^T \mathbf{w}_i - 1). \quad (19)$$

Tomando a sua derivada com relação à  $\mathbf{w}_i$  e igualando à 0, temos que:

$$\frac{\partial L(\mathbf{w}_i, \lambda)}{\partial \mathbf{w}_i} = 2\mathbf{C}_i \mathbf{w}_i + 2\alpha \mathbf{w}_i = \lambda \mathbf{1}^T = 0. \quad (20)$$

Podemos reescrever a equação acima da seguinte forma:

$$(\mathbf{C}_i + \alpha \mathbf{I}) \mathbf{w}_i = \frac{\lambda}{2} \mathbf{1}. \quad (21)$$

Agora temos uma equação fechada para encontrar  $\mathbf{w}_i$ , ou seja:

$$\mathbf{w}_i = \frac{\lambda}{2}(\mathbf{C}_i + \alpha \mathbf{I})^{-1} \mathbf{1}. \quad (22)$$

Podemos descartar, novamente, o termo  $\frac{\lambda}{2}$ , resultando no seguinte sistema linear:

$$(\mathbf{C}_i + \alpha \mathbf{I})\mathbf{w}_i = \mathbf{1}. \quad (23)$$

A diferença para a abordagem sem regularização é que estamos adicionando uma leve perturbação na diagonal principal da matriz  $\mathbf{C}_i$  por conta do termo  $\alpha \mathbf{I}$ .

Agora estamos no **subproblema 2**, ou seja, dado que temos os pesos de reconstrução, queremos saber quem são as coordenadas no espaço de menor dimensão que minimizam o erro quadrático (Equação 3). Este segundo problema de otimização possui duas restrições:

- 1 A média dos dados no espaço transformado  $\hat{\mathbf{X}}$  é zero, ou seja:

$$\hat{\boldsymbol{\mu}} = \frac{1}{m} \sum_{i=1}^m \hat{\mathbf{x}}_i = \mathbf{0}. \quad (24)$$

- 2 A matriz de covariância dos dados transformados é a matriz de identidade, ou seja, não existe correlação entre os componentes de  $\hat{\mathbf{x}} \in \mathbb{R}^d$ .



Ao contrário do passo anterior, aqui não conseguimos separar o problema em  $m$  subproblemas independentes. Da Equação 3, temos que:

$$\begin{aligned}\Phi(\mathbf{W}, \hat{\mathbf{X}}) &= \sum_{i=1}^m \left\| \hat{\mathbf{x}}_i - \sum_{\mathbf{x}_j \in \mathcal{N}(\mathbf{x}_i)} w_{ij} \hat{\mathbf{x}}_j \right\|^2 \\ &= \sum_{i=1}^m \left[ \left( \hat{\mathbf{x}}_i - \sum_{\mathbf{x}_j \in \mathcal{N}(\mathbf{x}_i)} w_{ij} \hat{\mathbf{x}}_j \right)^T \left( \hat{\mathbf{x}}_i - \sum_{\mathbf{x}_j \in \mathcal{N}(\mathbf{x}_i)} w_{ij} \hat{\mathbf{x}}_j \right) \right].\end{aligned}\quad (25)$$

Aplicando a distributiva na Equação 24, temos que:

$$\begin{aligned} \Phi(\mathbf{W}, \hat{\mathbf{X}}) = \sum_{i=1}^m & \left[ \hat{\mathbf{x}}_i^T \hat{\mathbf{x}}_i - \hat{\mathbf{x}}_i^T \sum_{\mathbf{x}_j \in \mathcal{N}(\mathbf{x}_i)} w_{ij} \hat{\mathbf{x}}_j - \left( \sum_j w_{ij} \hat{\mathbf{x}}_j \right)^T \hat{\mathbf{x}}_i + \right. \\ & \left. + \left( \sum_{\mathbf{x}_j \in \mathcal{N}(\mathbf{x}_i)} w_{ij} \hat{\mathbf{x}}_j \right)^T \left( \sum_{\mathbf{x}_j \in \mathcal{N}(\mathbf{x}_i)} w_{ij} \hat{\mathbf{x}}_j \right) \right]. \end{aligned} \quad (26)$$

Expandindo o somatório, temos que:

$$\begin{aligned}\Phi(\mathbf{W}, \hat{\mathbf{X}}) = & \sum_{i=1}^m \hat{\mathbf{x}}_i^T \hat{\mathbf{x}}_i - \sum_{i=1}^m \sum_{\mathbf{x}_j \in \mathcal{N}(\mathbf{x}_i)} w_{ij} \hat{\mathbf{x}}_i^T \hat{\mathbf{x}}_j - \sum_{i=1}^m \sum_{\mathbf{x}_j \in \mathcal{N}(\mathbf{x}_i)} w_{ji} \hat{\mathbf{x}}_j^T \hat{\mathbf{x}}_i + \\ & + \sum_{i=1}^m \sum_{\mathbf{x}_j \in \mathcal{N}(\mathbf{x}_i)} \sum_{\mathbf{x}_k \in \mathcal{N}(\mathbf{x}_i)} w_{ji} \hat{\mathbf{x}}_j^T w_{ik} \hat{\mathbf{x}}_k.\end{aligned}\tag{27}$$

Podemos reescrever a equação acima utilizando traços (soma dos elementos das diagonais) de matrizes, ou seja:

$$\begin{aligned}
\Phi(\mathbf{W}, \hat{\mathbf{X}}) &= \text{Tr}(\hat{\mathbf{X}}^T \hat{\mathbf{X}}) - \text{Tr}(\hat{\mathbf{X}}^T \mathbf{W} \hat{\mathbf{X}}) - \text{Tr}(\hat{\mathbf{X}}^T \mathbf{W}^T \hat{\mathbf{X}}) + \text{Tr}(\hat{\mathbf{X}}^T \mathbf{W}^T \mathbf{W} \hat{\mathbf{X}}) \\
&= \text{Tr}(\hat{\mathbf{X}}^T \hat{\mathbf{X}}) - \text{Tr}(\hat{\mathbf{X}}^T (\mathbf{W} \hat{\mathbf{X}})) - \text{Tr}((\mathbf{W} \hat{\mathbf{X}})^T \hat{\mathbf{X}}) + \text{Tr}((\mathbf{W} \hat{\mathbf{X}})^T (\mathbf{W} \hat{\mathbf{X}})) \\
&\hspace{20em} (28) \\
&= \text{Tr}(\hat{\mathbf{X}}^T \hat{\mathbf{X}} - \hat{\mathbf{X}}^T (\mathbf{W} \hat{\mathbf{X}}) - (\mathbf{W} \hat{\mathbf{X}})^T \hat{\mathbf{X}} + (\mathbf{W} \hat{\mathbf{X}})^T (\mathbf{W} \hat{\mathbf{X}})) \\
&= \text{Tr}(\hat{\mathbf{X}}^T (\hat{\mathbf{X}} - \mathbf{W} \hat{\mathbf{X}}) - (\mathbf{W} \hat{\mathbf{X}})^T (\hat{\mathbf{X}} - \mathbf{W} \hat{\mathbf{X}})) \\
&= \text{Tr}((\hat{\mathbf{X}} - \mathbf{W} \hat{\mathbf{X}})^T (\hat{\mathbf{X}} - \mathbf{W} \hat{\mathbf{X}})) \\
&= \text{Tr}(((\mathbf{I} - \mathbf{W}) \hat{\mathbf{X}})^T ((\mathbf{I} - \mathbf{W}) \hat{\mathbf{X}})) \\
&= \text{Tr}(\hat{\mathbf{X}}^T (\mathbf{I} - \mathbf{W}^T) (\mathbf{I} - \mathbf{W}) \hat{\mathbf{X}}).
\end{aligned}$$

Seja a matriz  $M \in \mathbb{R}^{m \times m}$  tal que:

$$M = (I - W)(I - W)^T, \quad (29)$$

em que  $W$  já encontramos anteriormente. Assim, temos o seguinte problema de otimização:

$$\begin{aligned} \hat{X}^* = \arg \min_{\hat{X}} \operatorname{Tr}(\hat{X}^T M \hat{X}) \\ \text{s.a. } \frac{1}{m} \hat{X}^T \hat{X} = 1. \end{aligned} \quad (30)$$

Como temos um problema de otimização com restrição, precisamos fazer uso da técnica dos multiplicadores de Lagrange.

O Lagrangiano da função é dado por:

$$L(\hat{\mathbf{X}}, \lambda) = \text{Tr}(\hat{\mathbf{X}}^T \mathbf{M} \hat{\mathbf{X}}) - \lambda \left( \frac{1}{m} \hat{\mathbf{X}}^T \hat{\mathbf{X}} - \mathbf{1} \right). \quad (31)$$

Derivando o Lagrangiano e igualando à zero, temos que:

$$2\mathbf{M}\hat{\mathbf{X}} - 2\frac{\lambda}{m}\hat{\mathbf{X}} = 0 \implies \mathbf{M}\hat{\mathbf{X}} = \beta\hat{\mathbf{X}}, \quad (32)$$

em que  $\beta = \frac{\lambda}{m}$ . Notem que, novamente, temos uma equação que envolve **autovalores** e **autovetores**. Neste caso,  $\hat{\mathbf{X}}$  são os autovetores da matriz  $\mathbf{M}$ . Como temos um problema de minimização (Equação 30), devemos escolher os  $d$  autovetores associados aos  $d$  menores autovalores. Entretanto, o menor autovalor é sempre 0 neste problema, e devemos descartar o autovetor associado à ele.

O algoritmo do LLE possui três parâmetros: número de dimensões do espaço reduzido  $d$ , tamanho da vizinhança  $k$  e parâmetro de regularização  $\alpha$  (em alguns casos). LLE é bastante sensível à escolha desses parâmetros:

- Caso o valor de  $k$  seja muito pequeno, não estaremos mapeando informações globais, e caso  $k$  seja muito grande a técnica perde sua característica não linear e tende a se comportar de maneira similar ao PCA.
- Caso  $\alpha$  seja escolhido de maneira inadequada, a decomposição espectral pode ser não funcionar corretamente (a ideia da regularização é tornar o processo possível em matrizes não inversíveis).
- Caso o valor de  $d$  seja muito pequeno, poderemos ter sobreposição de classes, e caso seja muito grande poderemos amplificar ruído.

Segue, abaixo, o algoritmo do LLE, cujas entradas são a matriz de dados  $\mathbf{X} \in \mathbb{R}^{m \times n}$ , o tamanho da vizinhança  $k$  e o número de dimensões de saída  $d$ . Note que  $\mathbf{x}_i$  representa a  $i$ -ésima linha de  $\mathbf{X}$ .

LLE( $\mathbf{X}, k, d$ )

Crie um grafo  $k$ -NN a partir de  $\mathbf{X}$ .

**for** each  $\mathbf{x}_i$  de  $\mathbf{X}$  **do**

    Calcule a matriz  $\mathbf{C}_i(j, k) = (\mathbf{x}_i - \mathbf{x}_j)^T(\mathbf{x}_i - \mathbf{x}_k)$ , tal que  $\mathbf{C}_i \in \mathbb{R}^{k \times k}$ .

    Resolva o sistema linear  $\mathbf{C}_i \mathbf{w}_i = \mathbf{1}$  para encontrar os pesos  $\mathbf{w}_i \in \mathbb{R}^k$ .

    Normalize os pesos  $\mathbf{w}_i$  para que  $\sum_j \mathbf{w}_{ij} = 1$ .

Construa a matriz  $\mathbf{W} \in \mathbb{R}^{m \times k}$  cujas linhas são os pesos estimados  $\mathbf{w}_i$ .

Calcule  $\mathbf{M} = (\mathbf{I} - \mathbf{W})^T(\mathbf{I} - \mathbf{W})$ , em que  $\mathbf{M} \in \mathbb{R}^{m \times m}$ .

Calcule os autovetores e autovalores da matriz  $\mathbf{M}$ .

Selecione os  $d$  autovetores não nulos associados aos  $d$  menores autovalores para definir a matriz  $\hat{\mathbf{X}} \in \mathbb{R}^{m \times d}$ , em que cada linha é um autovetor.