1. K-means. K均值算法.

Input : unlabeled data $D = \{\vec{x}^{(i)}\}_{i=1}^{N}$ $\vec{x}^{(i)} \in \mathbb{R}^{M}$.

Goal : find an assignment of points to clusters.

$$\vec{z} = \{z^{(1)}, z^{(2)}, \ldots, z^{(N)}\} \quad z^{(i)} \in \{1, \ldots k\}$$

K = # clusters is a hyperparameter.

cluster centers : $C = \{\vec{c}_1^{T}, \vec{c}_2^{T}, \ldots, \vec{c}_k^{T}\}^{T}$ $\vec{c}_j \in \mathbb{R}^{M}$.

Decision rule : assign each point $\vec{x}^{(i)}$ to its nearest center $\vec{c}_j$.

objective : $$C = \underset{C}{\arg\min} \sum_{i=1}^{N} \min_{j} \|\vec{x}^{(i)} - \vec{c}_j\|_2^2$$

$$= \underset{C}{\arg\min} \sum_{i=1}^{N} \min_{z^{(i)}} \|\vec{x}^{(i)} - \vec{c}_{z^{(i)}}\|_2^2.$$

$$C, \vec{z} = \underset{C, \vec{z}}{\arg\min} \sum_{i=1}^{N} \|\vec{x}^{(i)} - \vec{c}_{z^{(i)}}\|_2^2$$

$$= \underset{C, \vec{z}}{\arg\min} J(C, \vec{z}).$$

K-means in practice

Algo : ① Given $\vec{x}^{(1)}, \ldots, \vec{x}^{(N)}$.

② Initialize centers $C = \{\vec{c}_1, \ldots \vec{c}_k\}$

assignment z.

③ Repeat until convergence :

ⓐ $C = \underset{C}{\arg\min} J(C, \vec{z})$

ⓑ $z = \underset{z}{\arg\min} J(C, \vec{z})$

3ⓐ. $J(C, z) = \sum_{i=1}^{N} \| \vec{x}^{(i)} - \vec{C}_{z^{(i)}} \|_2^2$

$$= \sum_{j=1}^{k} \sum_{i: z^{(i)}=j} \| \vec{x}^{(i)} - \vec{C}_j \|_2^2$$

$$\vec{C}_1 = \underset{\vec{C}_1}{\operatorname{argmin}} \sum_{i: z^{(i)}=1} \| \vec{x}^{(i)} - \vec{C}_1 \|_2^2$$

$$\vdots$$

$$\vec{C}_k = \underset{\vec{C}_k}{\operatorname{argmin}} \sum_{i: z^{(k)}=1} \| \vec{x}^{(i)} - \vec{C}_k \|_2^2$$

$$\vec{C}_j = \frac{1}{N_j} \sum_{i: z^{(i)}=j} \vec{x}^{(i)}$$

3ⓑ: Find the closest cluster center $\vec{C}_j$ for each $\vec{x}^{(i)}$

$$z^{(i)} = \underset{j}{\operatorname{argmin}} \| \vec{x}^{(i)} - \vec{C}_j \|_2^2$$

2. DBSCAN.

基于密度的聚类 density-based clustering

邻域参数 (ε, MinPts).

核心对象.

密度直达. 密度可达.

3. 层次聚类.

Dendrogram. 树状图

1. Data for PCA.

What if your data is not centered? $\Rightarrow$ subtract off the sample mean. $(x^{(i)'} - \mu')$.

$x^{(i)} \in \mathbb{R}^M$

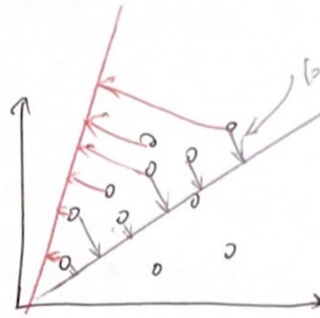
2. Strawman: Random Linear Projection. 随机投影

Algo: ① Random sample matrix. $V \in \mathbb{R}^{K \times M}$

② Project down $\underset{K\times 1}{\vec{v}^{(i)}} = \underset{K\times M}{V}\,\underset{M\times 1}{\vec{x}^{(i)}} \quad \forall i$

③. Reconstruct up $\tilde{x}^{(i)} = V^T \vec{v}^{(i)}$.

$\hookrightarrow$ loses information



PCA carefully construct V to preserve as much info as possible

3. Definition of PCA.

- Given K vectors. $\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_K$ where $\vec{v} \in \mathbb{R}^M$, the projection of a vector $x^{(i)} \in \mathbb{R}^M$ to a lower K-dimensional space is $\vec{v}^{(i)} \in \mathbb{R}^K$ where.

$$\vec{v}^{(i)} = \begin{bmatrix} \vec{v}_1^T \vec{x}^{(i)} \\ \vec{v}_2^T \vec{x}^{(i)} \\ \vdots \\ \vec{v}_K^T \vec{x}^{(i)} \end{bmatrix} = V\vec{x}^{(i)} \qquad \text{where } V = \begin{bmatrix} \vec{v}_1^T \\ \vec{v}_2^T \\ \vdots \\ \vec{v}_K^T \end{bmatrix}$$

- Def: PCA repeatedly chooses a next vector $\vec{v}_j$ that minimize the reconstruction error s.t $\vec{v}_j$ is orthogonal to $\vec{v}_1, \cdots \vec{v}_{j-1}$

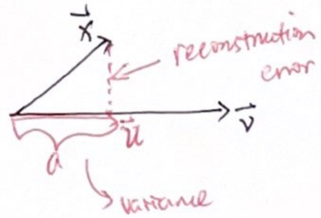
Recall $\vec{a}$ and $\vec{b}$ are orthogonal iff $\vec{a}^T b = 0$.

$\hookrightarrow$ K-dimensions in PCA are uncorrelated.

look the same from the other dimension

Any info. provided by one dimension is not provided by other dimensions.

Projection



length of projection of $\vec{x}$ onto $\vec{v}$

$$a = \vec{v}^T \vec{x} \quad \text{if } \|v\|_2 = 1$$
$$\frac{\vec{v}^T \vec{x}}{\|\vec{v}\|_2} \quad \text{otherwise}$$

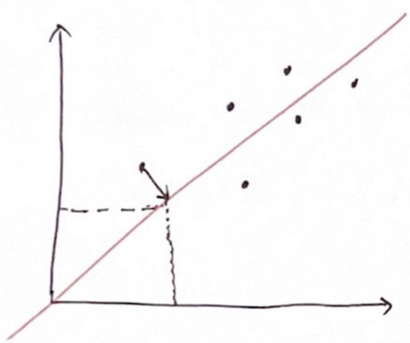
projection of $\vec{x}$ onto $\vec{v}$

$$\vec{u} = a\vec{v} = (\vec{v}^T \vec{x})\vec{v} \quad \text{if } \|v\|_2 = 1.$$

3. objective Function for PCA.

give the same $\vec{v}$. Equivalent.

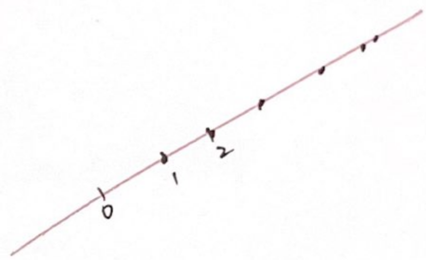
①. Minimize the Reconstruction error.



$$\vec{v} = \underset{\vec{v}}{\arg\min} \frac{1}{N} \sum_{i=1}^{N} \|\vec{x}^{(i)} - \underbrace{(\text{proj. of } x^{(i)})}_{\frac{\vec{v}^T \vec{x}^{(i)} \vec{v}}{\|v\|_2^2}}\|_2^2$$

$$= \underset{\vec{v}}{\arg\min} \frac{1}{N} \sum_{i=1}^{N} \|\vec{x}^{(i)} - (\vec{v}^T \vec{x}^{(i)})\vec{v}\|_2^2$$
$$\text{s.t } \|\vec{v}\|_2 = 1$$

②. Maximize the variance.



$$\vec{v} = \underset{\vec{v}}{\arg\max} \frac{1}{N} \sum_{i=1}^{N} \left(\text{projection length of } x^{(i)}\right)^2$$

$$= \underset{\vec{v}}{\arg\max} \frac{1}{N} \sum_{i=1}^{N} (\vec{v}^T \vec{x}^{(i)}) \quad \text{s.t } \|\vec{v}\|_2 = 1$$

$$= \underset{\vec{v}}{\arg\max} \frac{1}{N} \vec{v}^T (X^T X) \vec{v} \quad \text{s.t } \|v\|_2 = 1.$$

$$= \underset{\vec{v}}{\arg\max} \; \vec{v}^T \Sigma \vec{v}$$

$\Sigma = \frac{1}{N} X^T X$ covariance matrix

PCA Projections.

特征值

$$\vec{v}^{(i)} = \begin{bmatrix} \vec{v}_1^T \vec{x}^{(i)} \\ \vdots \\ \vec{v}_k^T \vec{x}^{(i)} \end{bmatrix}$$

$\vec{v}_1$ is the eigen vector w/1st largest eigenvalue.

$\vec{v}_2$ w/2nd

$\vdots$

$\vec{v}_k$ w/ k^{th}.

$$v_1 = \underset{v: \|\vec{v}\|^2 = 1}{\operatorname{argmax}} \; v^T \Sigma v.$$

$$L(v, \lambda) = v^T \Sigma v - \lambda(v^T v - 1).$$

$$\frac{\delta}{\delta v} L = 0 \Rightarrow \Sigma v - \lambda v = 0.$$

$$\Sigma v = \lambda v.$$

4. SVD. Singular value decomposition. 奇异值分解.

$A = U \Lambda V^T$ Λ is a diagonal matrix. 对角矩阵

U and V are orthogonal

$\hookrightarrow \Sigma = \frac{1}{N} V(\Lambda)^2 V^T.$ $(X = U \Lambda V^T)$.

$\therefore$ $X^T X$ and X share the same eigenvectors in their SVD.

$\therefore$ we can run SVD on X.