Transformada de Karhunen-Loève

Felipe Takaoka

ECA013 - f146024

fehtakaoka@gmail.com

5 de Setembro de 2018

Sumário

Sumário

Introdução e Motivação

- Baseada no teorema de Karhunen–Loève da representação de um processo estocástico como uma combinação linear de funções ortogonais;
- Base que minimiza o erro quadrático e dependente do processo;
- Também conhecida como Transformada de Hotelling ou Transformada dos Autovetores;
- Intimamente relacionado com a Análise em Componentes Principais (PCA);
- Largamente utilizado em processamento digital de sinais e imagens para compressão de dados (com perdas);

Sumário

Sumário

Definition

Uma matriz quadrada A é dita ortogonal se $A^{-1} = A^T$.

Theorem

Seja $A \in \mathbb{R}^{N \times N}$ uma matriz ortogonal. O conjunto dos seus vetores colunas $a_i \in \mathbb{R}^N$ forma uma base ortonormal (i.e. $\langle a_i, a_j \rangle = a_i^T a_j = \delta_{ij}$).

Definition

A transformada ortogonal por A de um sinal discreto $x \in \mathbb{R}^N$ é dada por:

$$y = A^T x$$

Ou seja, $y_i = a_i^T x = \langle a_i, x \rangle$. E, logo, a transformada inversa, $x = Ay$.

Teorema Espectral em Dimensão Finita

Se A é uma matriz real simétrica, então A é diagonalizável por uma matriz ortogonal.

Forma Canônica

$$A = Q\Lambda Q^{-1} = Q\Lambda Q^T$$

com $\Lambda = diag(\lambda_1, ..., \lambda_n)$ a matriz diagonal dos autovalores de A e $Q = [v_1, ..., v_n]$ a matriz ortogonal cujas colunas são compostas pelos autovetores de A

Sumário

Definition

Um vetor aleatório é um vetor cujos componentes são variáveis aleatórias (V.A.).

Sinais aleatórios

Um sinal físico pode ser considerado como um processo estocástico e uma amostragem produz, portanto, um vetor aleatório (discreto) $x = [x_1, x_2, ..., x_n]^T$.

Seja $x \in \mathbb{R}^{nx1}$ um vetor coluna aleatório real.

Média e Matriz de Variância(-Covariância) Teóricas

$$\mu := E(x) = [E(x_1), ..., E(x_n)]^T = [\mu_1, ..., \mu_n]^T$$

$$\Sigma := E[(x-\mu)(x-\mu)^T] = E(xx^T) - \mu\mu^T = \begin{bmatrix} \sigma_1^2 & \dots & \sigma_{1n}^2 \\ \vdots & \ddots & \vdots \\ \sigma_{n1}^2 & \dots & \sigma_n^2 \end{bmatrix}$$

Média e Matriz de Variância(-Covariância) Empíricas

Num total de n realizações, seja $x_i \in \mathbb{R}^{nx1}$ a i-ésima realização de um vetor coluna aleatório.

$$\hat{\mu} := \frac{1}{n}\sum_{k=1}^{n} x_k$$

$$\hat{\Sigma} := \frac{1}{n-1}\sum_{k=1}^{n} x_k x_k^T - \hat{\mu}\hat{\mu}^T$$

Obs.: se $\hat{\mu} = 0$ (facilmente obtido pela mudança de variável $x' = x - \hat{\mu}$)

$$\hat{\Sigma} = \frac{1}{n-1}\begin{bmatrix} | & & | \\ x_1 & \dots & x_n \\ | & & | \end{bmatrix}\begin{bmatrix} - & x_1^T & - \\ & \vdots & \\ - & x_n^T & - \end{bmatrix} = \frac{1}{n-1}(X^T X)$$

Transformada Ortogonal de um Vetor Aleatório

Momentos da transformada

Dados x um vetor aleatório e A uma matriz ortogonal, a transformada ortogonal $y = A^T x$ também é um vetor aleatório com:

$$\mu_y = E(x) = E(A^T x) = A^T E(x) = A^T \mu_x$$

$$\Sigma_y = E[(y - \mu_y)(y - \mu_y)^T] = ... = A^T \Sigma_x A$$

Sumário

Sumário

Transformada de Karhunen-Loève - Premissa

Premissa

Se $x = [x_1, x_2, ..., x_n]$ é um vetor aleatório que representa as N amostras de um sinal físico contínuo, as amostras próximas tendem a ser mais correlacionadas do que aquelas mais distantes.
Em outras palavras, dado x_i, pode-se prever x_{i+1} com muito mais "confiança" do que uma amostra x_j com $|j - i| >> 1$.

Sumário

Definição

Dado um vetor aleatório x, denotemos por ϕ_i o i-ésimo autovetor associado ao autovalor λ_i da matriz de variância de x, Σ_x.
Desta forma, pelo teorema espectral, como Σ_x é real e simétrica, seus autovetores formam uma base ortogonal. Assim, é ortogonal

$$\Phi := [\phi_1, ..., \phi_n]$$

Ou seja, $\Phi^{-1} = \Phi^T$. Além disso, se $\Lambda = diag(\lambda_1, ..., \lambda_n)$, tem-se:

$$\Phi^{-1}\Sigma_x\Phi = \Lambda$$

Transformada de Karhunen-Loève

$$y = \Phi^T x$$

Transformada Inversa

$$x = \Phi y$$

Sumário

Propriedades

É possível demonstrar que, dentre todas as transformações ortogonais, a KLT é ótima no sentido em que:

1. KLT "descorrelaciona" completamente o sinal;
2. KLT compacta de forma máxima a energia (informação) contida no sinal;

Propriedade 1 - KLT "descorrelaciona" o sinal

Momentos da Transformada

Média:

$$\mu_y = \Phi^T \mu_x$$

Matriz de variância:

$$\Sigma_y = ... = \Phi^T \Sigma_x \Phi = \Lambda = diag(\lambda_1, ..., \lambda_n)$$

$$\Sigma_y = \begin{bmatrix} \sigma_1^2 & \dots & \sigma_{1n}^2 \\ \vdots & \ddots & \vdots \\ \sigma_{n1}^2 & \dots & \sigma_n^2 \end{bmatrix} = \begin{bmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_n \end{bmatrix} = \Lambda$$

- Depois da transformada, a matriz de covariância do sinal é diagonalizada, ou seja, o sinal é "descorrelacionado";
- $\lambda_i = \sigma_i^2$

Propriedade 2 - KLT compacta a energia

- A KLT é a transformação ortogonal que melhor "redistribui" a energia/informação contida no sinal de forma que ela seja compactada num menor número de componentes;
- Define-se a energia contida nos primeiros $m < n$ componentes pela função:

$$S_m(A) := \sum_{i=1}^{m} E\left[(y_i - \mu_{y_i})^2\right] = \sum_{i=1}^{m} \sigma_{y_i}^2$$

Propriedade 2 - KLT compacta a energia

- Para a transformação KLT:

$$S_m(\Phi) = \sum_{i=1}^{m} \lambda_i = \sum_{i=1}^{m} \sigma_{x_i}^2$$

- Desta forma, se ordenarmos os autovalores de Σ_x em ordem decrescente, i.e. $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m \geq \cdots \geq \lambda_n$, então estaremos maximizando $S_m(\Phi)$;

Sumário

Construção Geral

Tendo em mente as propriedades da KLT, podemos utilizá-la para a compressão de dados de um sinal x, fazendo-se uma redução de dimensionalidade na sua representação da seguinte maneira:

1. Encontre a média m_x e a matriz de covariância Σ_x;
2. Encontre os autovetores λ_i de Σ_x ordenados de forma decrescente e os seus autovalores correspondentes;
3. Fixe uma porcentagem de perda de informação admissível (equiv. porcentagem mínima de informação contida do sinal original), e.g. 95%, e escolha o menor $m < n$ tal que:

$$\frac{\sum_{i=1}^{m} \lambda_i}{\sum_{i=1}^{n} \lambda_i} = \frac{\sum_{i=1}^{m} \sigma_{x_i}^2}{\sum_{i=1}^{n} \sigma_{x_i}} \geq 95\%$$

4. Construa a matriz de transformação nxm composta pelos autovetores ordenados de acordo com os seus respectivos autovalores:

$$\Phi_m = [\phi_1, \dots, \phi_m]_{nxm}$$

Construção Geral

5. Aplique a transformação ortogonal KLT:

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}_{mx1} = \Phi_m^T x = \begin{bmatrix} - & \phi_1^T & - \\ & \vdots & \\ - & \phi_m^T & - \end{bmatrix}_{mxn} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}_{nx1}$$

6. Reconstrua x aplicando a transformada KLT inversa:

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}_{nx1} = \Phi_m y = \begin{bmatrix} | & & | \\ \phi_1 & \dots & \phi_m \\ | & & | \end{bmatrix}_{nxm} \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}_{mx1}$$

Construção para Imagens

Dado N imagens de dimensão (r, c) com $K = rc$ pixels cada uma, seja $x_{c_i} \in \mathbb{R}^K$ o vetor coluna cujos componentes representam os pixels da imagem i. Desta maneira, construimos a matriz $X \in \mathbb{R}^{KxN}$ concatenando os vetores colunas de cada imagem:

$$X_{KxN} = \begin{bmatrix} | & & | \\ x_{c_1} & \dots & x_{c_N} \\ | & & | \end{bmatrix} = \begin{bmatrix} - & x_{r_1}^T & - \\ & \vdots & \\ - & x_{r_K}^T & - \end{bmatrix}$$

Assim, $x_{r_i} \in \mathbb{R}^N$ é o vetor coluna cujas componentes representam os valores dos pixels na mesma posição dos N diferentes canais/imagens.

Construção para Imagens

Assume-se, sem perda de generalidade, que os vetores $x_{c_i}, \forall i \in \{1, \ldots, K\}$ e $x_{r_j}, \forall j \in \{1, \ldots, N\}$ têm média nula. Desta maneira, a KLT pode ser aplicada de duas formas diferentes a depender do que é tratado como realização do evento aleatório:

K realizações do vetor $x_{r_i} \in \mathbb{R}^N$

Dada uma imagem i, os valores de seus pixels em todas as coordenadas são as realizações.

N realizações do vetor $x_{c_j} \in \mathbb{R}^K$

Dada uma localização de um pixel, os valores de seus pixels nos diferentes canais/imagens são as realizações.

Construção para Imagens

K realizações do vetor $x_{r_i} \in \mathbb{R}^N$

$$\Sigma_r = \frac{1}{K-1}(X^T X)$$

Se U denota a matriz ortogonal dos autovetores ordenados, a transformada KLT pode ser aplicada a cada um dos K vetores linhas de X:

$$y_{r_i} = U^T x_{r_i} \implies Y^T = U^T X^T \iff Y = XU$$

N realizações do vetor $x_{c_j} \in \mathbb{R}^K$

$$\Sigma_c = \frac{1}{N-1}(X X^T)$$

Se V denota a matriz ortogonal dos autovetores ordenados, a transformada KLT pode ser aplicada a cada um dos N vetores colunas de X:

$$z_{c_j} = V^T x_{c_j} \implies Z = V^T X$$

Construção para Imagens

Theorem

Os problemas dos autovalores associado às duas matrizes de covariância Σ_r e Σ_c são equivalentes no sentido em que produzem o mesmo conjunto de autovalores e seus autovetores associados satisfazem a relaçao $V = XU$.

Corolário

A escolha da resolução do problema do autovalor pode ser feita para aquela com menor dimensão

Sumário

Aplicações - Reconhecimento de Imagens

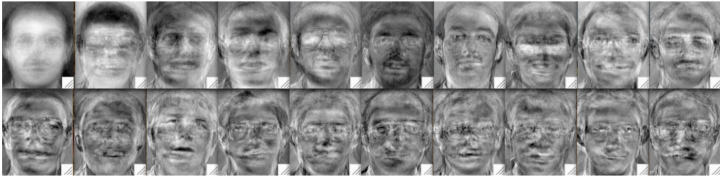
- Tendo em mente que não é necessário representar fielmente uma imagem pelo seus K pixels, podemos reduzir sua dimensionalidade de forma a diminuir seu espaço de armazenamento. Consequentemente, algoritmos de classificação de objetos em imagens, como redes neurais, se beneficiam amplamente desta representação em um espaço de dimensão reduzida;

Aplicações - Reconhecimento de Imagens

Eigenfaces and face recognition

Twenty images of faces ($N = 20$):

The eigen-images after KLT:



Aplicações - Reconhecimento de Imagens

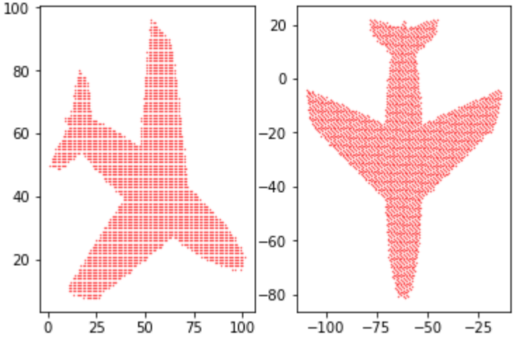
Percentage of energy contained in the

components	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
percentage energy	48.5	11.6	6.1	4.6	3.8	3.7	2.6	2.5	1.9	1.9	1.8	1.6	1.5	1.4	1.3	1.2	1.1	1.1	0.9	0.8
accumulative energy	48.5	60.1	66.2	70.8	74.6	78.3	81.0	83.5	85.4	87.3	89.	90.7	92.2	93.6	94.9	96.1	97.2	98.2	99.2	100.0

Reconstructed faces using 95% of the total information (15 out of 20 components):

Aplicações - Alinhamento de objetos

- Para imagens binárias, é possível alinhar objetos bidimensionais ao longo do seu autovetor principal (associado ao maior autovalor), o que pode ser importante para reconhecimento de objetos;



Sumário

Referências

- http://fourier.eng.hmc.edu/e161/lectures/klt/klt.html
- http://www.dca.fee.unicamp.br/~lotufo/cursos/ia-636-2000/labs/5rsoares/lab12/index.html
- https://en.wikipedia.org/wiki/Karhunen-Loeve_theorem
- https://pt.wikipedia.org/wiki/Matriz_ortogonal#cite_note-7
- https://en.wikipedia.org/wiki/Symmetric_matrix
- ROUVIÈRE, L. Statistique - Notes de Cours. CentraleSupélec, 2016.
- RICHARD, P.Y. Représentation et analyse statistique des signaux - Notes de Cours. CentraleSupélec, 2016.