

EE269

Signal Processing for Machine Learning

Lecture 18

Instructor : Mert Pilanci

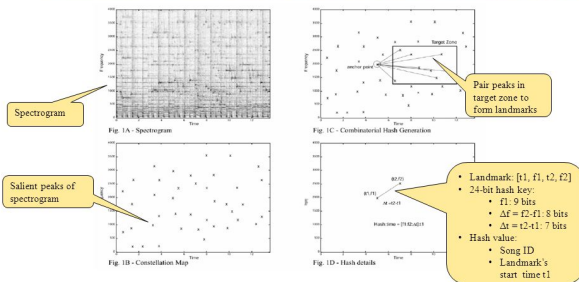
Stanford University

March 13, 2019

Application: Audio Fingerprinting

- ▶ Spectrogram is reduced to prominent time-frequency locations
- ▶ Constellations are robust against noise and outliers

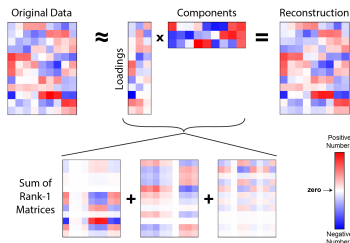
Shazam: Landmarks as Features



(Avery Wang, 2003)



Matrix Decompositions



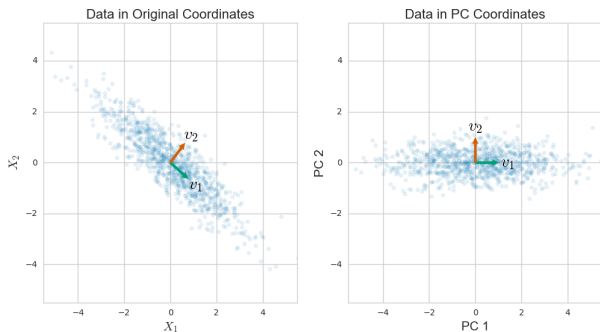
$X : n \times d$ data matrix

factors $W : n \times k$, and $Z : n \times k$

$$\min_{W,Z} ||X - WZ^T||_F^2$$

- ▶ can be solved by Singular Value Decomposition $X = U\Sigma V^T$
- ▶ optimal $WZ^T = U_{1:k}\Sigma_{1:k}V_{1:k}^T$
- ▶ Principal Component Analysis basis: $V_{1:k}$

Matrix Decompositions



$X : n \times d$ data matrix

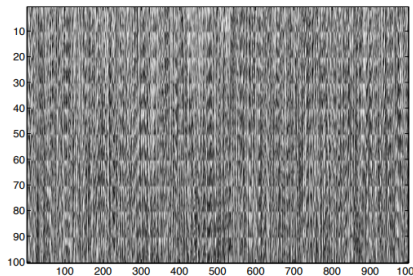
$$\min_{W,Z} \|X - WZ^T\|_F^2$$

- ▶ can be solved by Singular Value Decomposition $X = U\Sigma V^T$
- ▶ optimal $WZ^T = U_{1:k}\Sigma_{1:k}V_{1:k}^T$
- ▶ Principal Component Analysis basis: $V_{1:k}$ top k eigenvectors of $X^T X = V\Sigma^2 V^T$

Principal Component Analysis

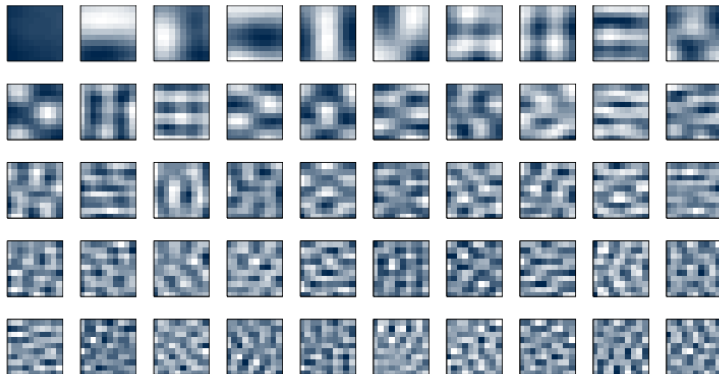


$\text{vec}(\square)$



slide credit: P Smaragdis

Principal Component Analysis



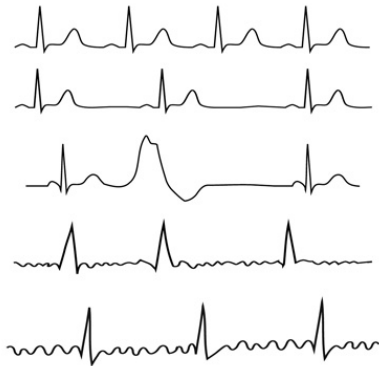
slide credit: P Smaragdis

Application: Arrhythmia Detection

Normal heart rhythm



Irregular heart rhythm



Application: Arrhythmia Detection

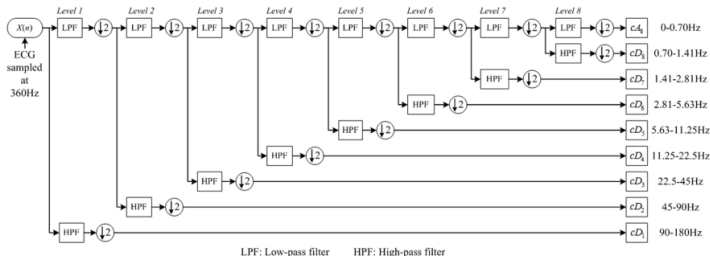


Figure 3. The decomposition process of the 8-level WMRA.

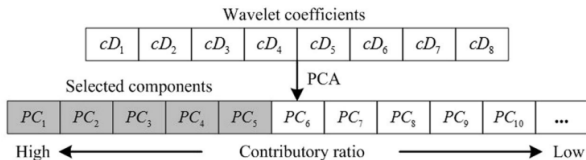


Figure 4. Low-dimensional feature vector generated by PCA using wavelet coefficients.

Application: Arrhythmia Detection

Literatures and feature extraction methods	Feature selection (dimension)	Beat types	Training/test beats	Classifiers	Independent training/test data	k-fold cross validation	SEN (%)	SPE (%)	ACC (%)
Spectral correlation ¹	Yes (88)	5	Totally 6259	SVM	Unknown	10-fold	99.20	99.70	98.60
Wavelet transform, morphological features ⁵	No (28)	5	10675/93894	Artificial neural network	No	No	88.60	96.18	97.86
Morphological features ⁷	Yes (6)	6	35848/35848	Linear discriminant analysis	No	No	91.19	98.65	94.03
Morphological features ⁸	No (13)	3	600/30273	SVM, neural network	No	No	98.52	99.19	97.14
Time domain features ⁹	No (9)	6	42427/14142	Decision tree	No	No	97.50	99.80	99.51
Morphological features ¹⁰	No (16)	3	15509/8081	SVM, neural network	Yes	No	92.82	93.74	92.85
Morphological features ¹¹	No (8)	5	12570/12570	Regression neural network	No	No	85.50	99.40	99.40
Fourier transform, wavelet package ¹⁴	Yes (70)	16	3345/2542	k-NN	No	No	85.59	99.56	93.59
Wavelet transform, cosine transform ¹⁵	Yes (18)	4	720/360	SVM	Unknown	No	98.60	95.50	96.50
Wavelet transform ¹⁶	Yes (24)	5	900/900	SVM, genetic algorithm	No	No	98.50	99.69	98.80
Higher order spectral ¹⁷	No (7)	5	330/500	SVM	Unknown	No	90.00	87.93	85.79
Wavelet transform ¹⁸	Yes (20)	4	360/360	SVM	Unknown	No	98.62	99.54	98.61
Temporal and spectral features ²¹	Yes (15)	6	1440/720	SVM	No	No	97.60	93.80	95.20
Temporal and spectral features ²²	Yes (13)	8	Totally 17857	SVM	No	5-fold	95.00	99.00	98.60
Higher order statistics, wavelet packet ²⁷	Yes (28)	5	3345/2542	k-NN	Yes	No	89.80	97.80	—
Hilbert-Huang transform ³¹	Yes (18)	6	10700/10700	SVM	No	No	98.64	99.77	99.51
Wavelet transform ⁴⁶	Yes (18)	5	Totally 101352	SVM	Yes	44-fold	—	—	86.40
		16	24100/86009		No	99.32	—	—	99.01
Approximate entropy, wavelet packet ⁴⁷	Yes (9)	5	145/145	SVM, PNN	Unknown	No	98.70	99.70	98.60
Non-linear and center-clipping transform ⁴⁸	No (5)	5	13640/13640	Wavelet neural network	No	No	98.78	99.70	98.78
Eigenvector method ⁴⁹	Yes (12)	4	360/360	Recurrent neural network	Unknown	No	98.89	99.25	98.06
Higher order statistics ⁵⁰	No (24)	5	4000/14299	RBF neural network	No	No	92.93	98.52	95.18
Geometrical features ⁵¹	No (18)	7	4035/3150	SVM, k-NN, BPNN	No	No	97.52	99.65	98.06
Wavelet transform, morphological features ⁵²	Yes (8)	3	50928/49636	Linear discriminant analysis	Yes	No	80.00	—	94.00
Wavelet transform, linear prediction model ⁵³	No (12)	3	50554/49273	Linear discriminant analysis	Unknown	No	86.50	—	86.50
Cross correlation ⁵⁴	No (30)	3	41961/51285	Artificial neural network	Unknown	No	97.49	—	95.24
WMRA [This work]	Yes (12)	6	Totally 107049	SVM	Yes	10-fold	44.40	88.88	81.47
					No		99.09	99.82	99.70

Nonnegative Matrix Factorization

$X : n \times d$ data matrix

$$\min_{W \geq 0, Z \geq 0} \|X - WZ\|_F^2$$

- ▶ non-negative data
- ▶ pixel values, energies, counts are non-negative valued

Nonnegative Matrix Factorization



- PCA and Fisher's LDA basis vectors are hard to interpret

Nonnegative Matrix Factorization

$X : n \times d$ data matrix

$$\min_{W \geq 0, Z \geq 0} \|X - WZ\|_F^2$$

- ▶ No closed form solution!
- ▶ Solution is not unique: $X = WDD^{-1}Z = WZ$

Nonnegative Matrix Factorization

$X : n \times d$ data matrix

$$\min_{W \geq 0, Z \geq 0} \|X - WZ\|_F^2$$

- ▶ No closed form solution!
- ▶ Solution is not unique: $X = WDD^{-1}Z = WZ$
- ▶ Heuristic algorithm:

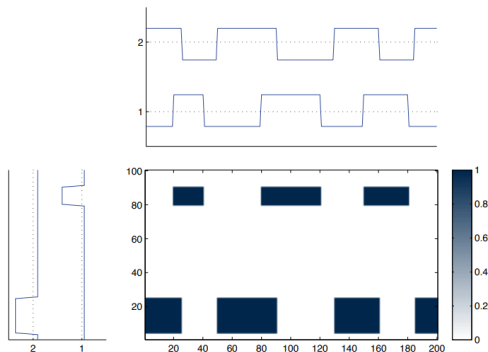
$$W \leftarrow \arg \min_W \|X - WZ\|_F = XZ^\dagger, \quad W \leftarrow (W)_+$$

$$Z \leftarrow \arg \min_Z \|X - WZ\|_F = W^\dagger X, \quad Z \leftarrow (Z)_+$$

iterate

Nonnegative Matrix Factorization

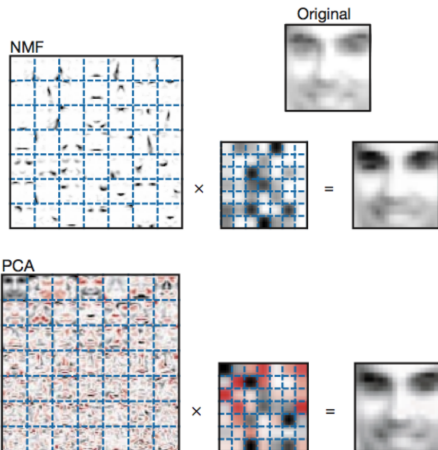
$$\min_{W \geq 0, Z \geq 0} \|X - WZ\|_F^2 \quad \text{implies} \quad X \approx WZ$$



- Spectrogram model: columns of W contain frequencies (e.g. musical notes), rows of Z contain time information

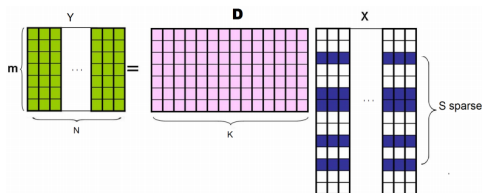
Nonnegative Matrix Factorization for Images

$$\min_{W \geq 0, Z \geq 0} \|X - WZ\|_F^2 \quad \text{implies} \quad X \approx WZ$$



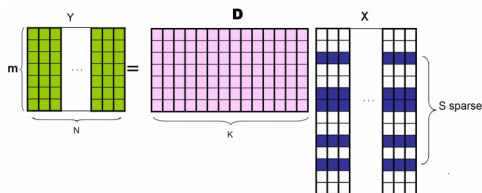
Dictionary Learning

$$\min_{x_1, \dots, x_n, D} \sum_{i=1}^n \|Dx_i - y_i\|_2^2 \quad \text{s.t.} \quad \|x_i\|_0 \leq s$$



Dictionary Learning

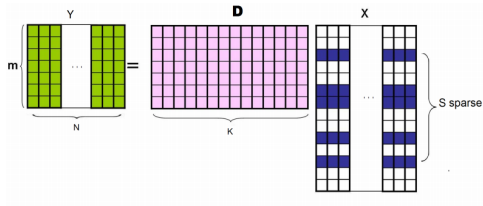
$$\min_{x_1, \dots, x_n, D} \sum_{i=1}^n \|Dx_i - y_i\|_2^2 \quad \text{s.t.} \quad \|x_i\|_0 \leq s$$



- ▶ Matrix factorization interpretation:
 $DX \approx Y$ where X has sparse columns

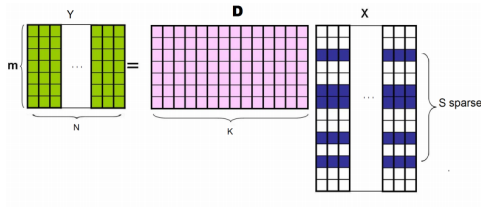
Dictionary Learning : ℓ_1 relaxation

$$\min_{x_1, \dots, x_n, D} \sum_{i=1}^n \|Dx_i - y_i\|_2^2 + \lambda \|x_i\|_1$$



Dictionary Learning : ℓ_1 relaxation

$$\min_{x_1, \dots, x_n, D} \sum_{i=1}^n \|Dx_i - y_i\|_2^2 + \lambda \|x_i\|_1$$



► heuristic algorithm

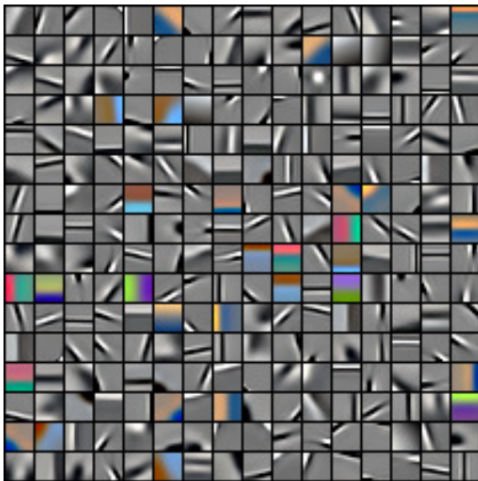
minimize over x_i : ℓ_1 regularized Least Squares problem

gradient descent step over D :

$$D \leftarrow D - \mu \sum_i (Dx_i - y_i) x_i^T$$

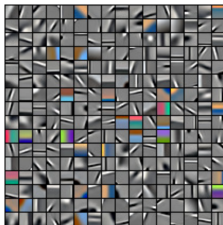
Learned Dictionaries from Images

$$\min_{x_1, \dots, x_n, D} \sum_{i=1}^n \|Dx_i - y_i\|_2^2 + \lambda \|x_i\|_1$$

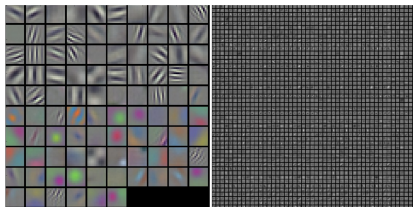


Dictionary Learning vs Deep Learning

- ▶ Dictionary learning

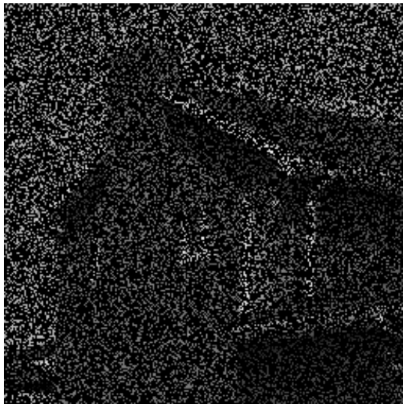


- ▶ Alexnet: Convolutional net for image classification



Dictionary Learning Applications

[Mairal, Sapiro, and Elad, 2008d]



Dictionary Learning Applications

Inpainting, [Mairal, Elad, and Sapiro, 2008b]



Dictionary Learning Applications

Inpainting, [Mairal, Elad, and Sapiro, 2008b]



Other SP/ML methods

- ▶ Independent Component Analysis
- ▶ Autoencoders
- ▶ Recurrent Neural Nets
- ▶ Manifold learning methods (e.g., Laplacian Eigenmaps, Isomap)
- ▶ Tensor Decompositions (e.g. Tensor Principal Component Analysis)
- ▶ Latent and Mixture Models (Hidden Markov Models, Low rank models)
- ▶ Bayesian Methods
- ▶ State Space Models and tracking (e.g., Kalman filter)