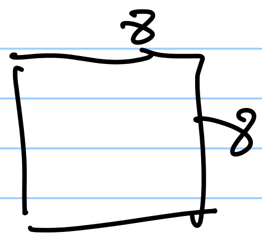
(1) Lecture 3. Stat 238. Winter 2015

Note Title 1/14/2015

What can happen in an 8x8 image window?



Theoretically 256^{64} possible images
But which ones happen?

How to represent images?

- Basis Functions / Fourier Series
- Overcomplete bases, sparse coding
- Learning bases : (i) PCA, (ii) Sparsity, (iii) Matched Filters
- Shift invariance - Mini-epitomes, Active Patches

(2)

Representing Images in terms of basis function.

Classic: Orthogonal set of basis functions

$$\{ b_i(x) : i = 1 \text{ to } N \},$$

$$\sum_x \left(b_i(x) \right)^2 = 1$$

or $$\int dx \left(b_i(x) \right)^2 = 1$$

$$\sum_x b_i(x)\, b_j(x) = 0, \text{ if } i \neq j$$

$$\int dx\, b_i(x) b_j(x) = 0, \; i \neq j$$

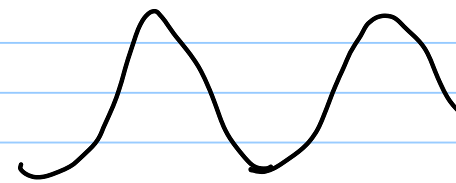
2D 8x8 patch

Examples:

- Sinusoids / Fourier Analysis
- Haar Bases
- Impulse Function.

(3) Jpeg coding.

Choose basis functions to be sinusoids



Represent image by $I(x) = \sum_i \alpha_i b_i(x)$

because the bases are orthonormal, we can solve to get $\alpha_i = \sum_x I(x) b_i(x)$ (or $\int dx \ldots$)

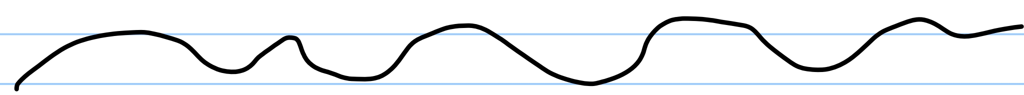
Image represented by the coefficients $\{\alpha_i\}$

Also we could minimize an error $\sum_x \{I(x) - \sum_i \alpha_i b_i(x)\}^2$

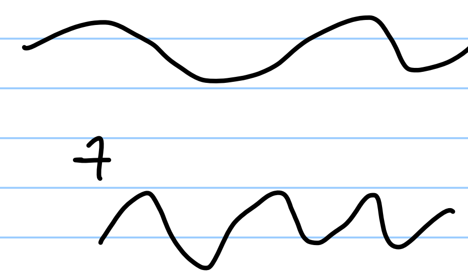
and try to restrict the no. of non-zero α's. | This gives standard image format Jpeg.

(4) Sinusoids / Fourier Theory work well if the image can be approximated well by a set of sinusoids.

E.g.



=



But an image like this

is better approximated by a set of impulse functions

And an image like this

is badly modeled by either.

(5) Over-Complete Bases.

Represent the image by an over-complete set
E.g. all the sinusoids and all the bases.

But now we have a problem.

There will be many ways to represent the image in form

$$I(x) = \sum_i \alpha_i b_i(x)$$

(because we could represent it by sinusoids only, or by impulse functions only, or by combinations.)

(6) Sparsity. L1 - Sparsity

regularizer

L1 norm.

Determine the α's by minimizing

$$E[\alpha] = \sum_x \left\{ I(x) - \sum_i \alpha_i b_i(x) \right\}^2 + \lambda \sum_i |\alpha_i|$$

pays a penalty for the coefficients. (α's)

Note: $E[\alpha]$ is a convex function.

There are efficient algorithms to estimate $\hat{\alpha} = \arg\min E[\alpha]$

Solution $\hat{I}(x) = \sum_i \hat{\alpha}_i b_i(x)$.

By a "miracle" (later in course) many of the α's will be zero.

(7) <u>Extreme Sparsity</u> L-0 sparsity

Set of basis functions $\{b_i(x)\}$,

represent each image by one basis function only

$$E[\alpha] = \sum_x \left[I(x) - \sum_i \alpha_i b_i(x) \right]^2$$ with constraint.

only one $\alpha_i \neq 0$. (recall that $\sum_x \{b_i(x)\}^2 = 1$)

Algorithm to estimate $\hat{\alpha} = \arg\min_\alpha E[\alpha]$.

Set $\hat{\alpha}_i = \arg\min_{\alpha_i} \sum_x [I(x) - \alpha_i b_i(x)]^2 = \sum_x I(x) b_i(x)$

choose $\hat{i} = \min_i \sum_x [I(x) - \hat{\alpha}_i b_i(x)]^2 \rightarrow$ set $\alpha_{\hat{i}} = \hat{\alpha}_{\hat{i}}$, $\alpha_j = 0$, otherwise

(8) Comments.

We described three ways to represent images using basis functions.

sparser

Classical: eg. Fourier Theory. / Haar Basis Function

L1 - Sparsity

L0 . Sparsity

Both overcomplete.

But what bases to use?

We can use the bases (20th century maths)

or we can learn them from data (21th century)

(9) Learning the bases.

Lets start with the classical approach.

Bases are orthogonal : $\sum_x b_i(x) b_j(x) = \delta_{ij}$ $= 1$ if $i = j$, $= 0$ if $i \neq j$

$\times$ Kronecker Delta.

Dataset of Images $\{ I^\mu(x) : \mu \in \Lambda \}$

Energy Function $E[b, \alpha] = \frac{1}{|\Lambda|} \sum_{\mu \in \Lambda} \sum_x \{ I^\mu(x) - \sum_i \alpha_i^\mu b_i(x) \}^2$

Note: basis functions are the same for all images
the coefficients α_i^μ vary between images

[16]

Minimize

$$E[b, \alpha] = \frac{1}{|\Lambda|} \sum_{\mu \in \Lambda} \sum_{x} \left\{ I^{\mu}(x) - \sum_{i} \alpha_i^{\mu} b_i(x) \right\}^2$$

w.r.t. $\{b, \alpha\}$.

This is simply Principal Component Analysis (PCA)

Provided we extract the means from the images

$$I^{\mu}(x) \to I^{\mu}(x) - \frac{1}{\Lambda} \sum_{\mu \in \Lambda} I^{\mu}(x)$$, so that $\sum_{\mu} I^{\mu}(x) = 0$

(after subtraction)

(I 1) Solution.

The basis function $b_i(x)$ are the eigenvectors of the correlation matrix $K(x,y) = \frac{1}{|\Lambda|} \sum_{\mu \in \Lambda} I^{\mu}(x) I^{\mu}(y)$

The coefficients $\alpha_i^{\mu} = \sum_x b_i(x) I^{\mu}(x)$

(as before).

We can restrict the number of basis functions by only uses those eigenvectors whose eigenvalues are above a threshold T.

$\sum_y K(x,y)\, b_i(y) = \lambda_i b_i(x)$ ✓ keep $b_i(x)$ if $\lambda_i > T$

(12) What are the eigenvectors of image patches?

Claim. If the image patches are randomly drawn from real images, then the eigenvectors are sinusoids?

Why? Because images are shift-invariant.

$$K(x,y) = F(x-y)$$

The correlation function depends only on the different $x-y$

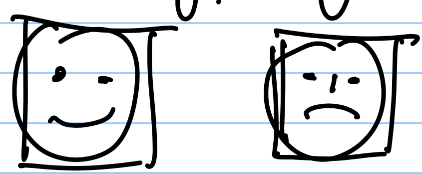
Eigenvectors:

$$\sum_y F(x-y)e(y) = \lambda e(x)$$

sinusids → proof - apply the convolution theorem

(13) So PCA doesn't help much. You know you will get sinusoids before you look at the images..

It is different if we align the images. For example if we have images of faces and center them in the image patch.



The alignment means that we remove shift-invariance.

But it is not possible to align general images.

(14) Now try sparsity — Olshausen & Field. 1996

$$E[b, \alpha] = \frac{1}{|\Lambda|} \sum_{\mu \in \Lambda} \sum_x \Big(I(x) - \sum_i \alpha_i^\mu b_i(x) \Big)^2 + \lambda \sum_{\mu \in \Lambda} \sum_i |\alpha_i^\mu| \, ,$$

Constraint. $\sum_x (b_i(x))^2 = 1$

Minimize w.r.t. b & α.

Note: $E[b, \alpha]$ is convex in α if b is fixed (sparsity)

It is convex in b if α is fixed.

Alternating Algorithm:
- Initialize b's
- Minimize w.r.t. α and b alternatively
- Guaranteed to converge to local minima.

code available online

(15) <u>Olshausen & Field</u> (paper on website)

applied this to natural images. <u>See examples.</u>

This gives more interesting bases <u>than</u> PCA.

<u>Note</u>: Deep Neural Networks obtain similar bases.

(16) <u>Final Alternative</u>   L0 sparsity. $\sum_x \langle b_i(x)\rangle_i^2 = 1$

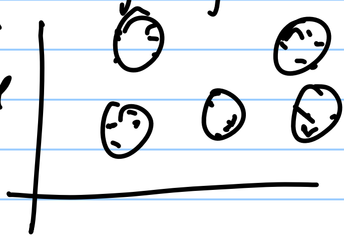
<u>Minimize</u> $E[b,\alpha] = \frac{1}{|\Lambda|}\sum_{\mu\in\Lambda}\sum_x \left\{ I^\mu(x) - \sum_i \alpha_i^\mu b_i(x)\right\}^2$

with constraint, that only one α_i^μ is non-zero for each μ.

How to minimize?

Convert this to <u>k-means</u> clustering. Requires normalizing each image $I^\mu(x) \to \frac{I^\mu(x)}{\sqrt{\sum_x \langle I^\mu(x)\rangle_i^2}}$ so that $\sum_x \langle I^\mu(x)\rangle_i^2 = 1$, implies that best $\alpha_i^\mu = 1$.

parameter space



k-means.

(17) Extensions:

All the previous methods have problems with shift-invariance.

The basis functions are encoding the shifts as well as the image patterns → see proe point.

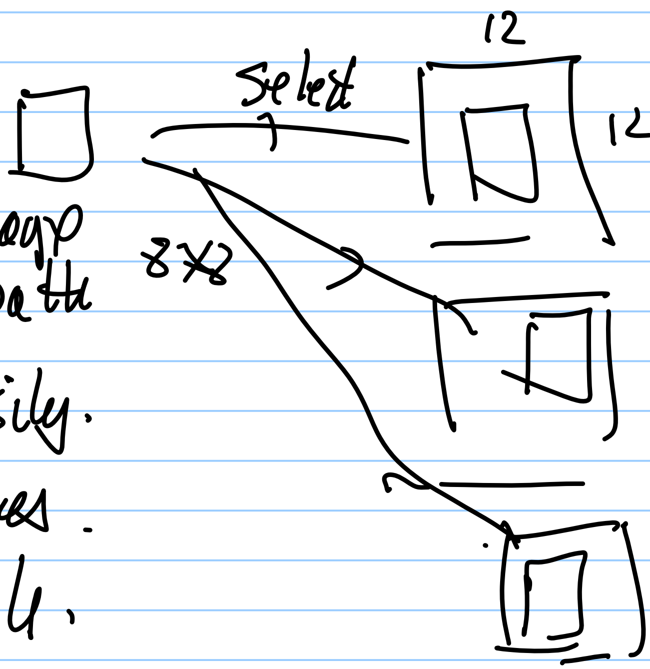
One solution → Mini-Epitomes. G. Papandreou, L-C. Chen, A.L. Yuille 2014

(18)

Extensions.

Mini-Epitomes

mini-epitomes.



This is like an extension of L0 sparsity. But with smarter patches. See notes on powerpoint.

Can be learnt by the EM algorithm – extending k-means.

(19)

Extensions

One result: A small set of mini-epitomes. 128 is able to represent most image patches in 10,000 images with good accuracy. So the no. of possible image patches may not be too enormous.

Another approach Active Patches. J. Mao, J. Zhu, A. Yuille 2014

Allow the patch to be deformed when it matches the image - see powerpoints

(20)

Why Image Patches?

Helps capture what locally happens in images. Can rediscover edges by examining the bases learnt from images (by L_0, or mini-epitomes).

- Can be used for image processing applications:
  (i) image denoising, (ii) super-resolution. (state of the art).
- Can be used for high-level vision tasks. (later in course)