CSE 559A: Computer Vision

Fall 2018: T-R: 11:30-1pm @ Lopata 101

Instructor: Ayan Chakrabarti (ayan@wustl.edu).
Course Staff: Zhihao Xia, Charlie Wu, Han Liu

http://www.cse.wustl.edu/~ayan/courses/cse559a/

Sep 13, 2018

ADMINISTRIVIA

- Tomorrow
  - Zhihao's Office Hours back in Jolley 309: 10:30am-Noon

- This Friday: Regular Office Hours
- Next Friday: Recitation for PSET 1
  - Try all problems before coming to recitation
- Monday Office Hours again in Jolley 217

IMAGE RESTORATION

Statistics / Estimation Recap

Setting

- There is a true image (or image-like object) x that we want to estimate
- What we have is some degraded observation y that is based on x
- The degradation might be "stochastic": so we say $p(y|x)$

Simplest Case (single pixel: $x, y \in \mathbb{R}$):

$$y = x + \sigma\epsilon \rightarrow p(y|x) = \mathcal{N}(y; \mu = x, \sigma^2)$$

Estimate x from y: "denoising"

IMAGE RESTORATION

Statistics / Estimation Recap

- $p(y|x)$ is our observation model, called the likelihood function
  - Likelihood of observation y given true image x
- Maximum Likelihood Estimator: $\hat{x} = \arg\max_x p(y|x)$

For $p(y|x) = \mathcal{N}(y; x, \sigma^2)$, what is ML Estimate ?

$$p(y|x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y-x)^2}{2\sigma^2}\right)$$

$\hat{x} = y$

That's a little disappointing.

IMAGE RESTORATION

Statistics / Estimation Recap

- What we really want to do is maximize $p(x|y)$, not the likelihood $p(y|x)$
  - $p(y|x)$ is the distribution of observation y given true image x
  - $p(x|y)$ is the distribution of true x given observation y
- Bayes Rule

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}$$

Simple Derivation

$$p(x,y) = p(x)p(y|x) = p(y)p(x|y) \rightarrow p(x|y) = \frac{p(y|x)p(x)}{p(y)}$$

$$p(y) = \int p(y|x')p(x')dx'$$

IMAGE RESTORATION

Statistics / Estimation Recap

$$p(x|y) = \frac{p(y|x)p(x)}{\int p(y|x')p(x')dx'}$$

- $p(y|x)$: Likelihood
- $p(x)$: "Prior" Distribution on x
  - What we believe about x *before* we see the observation
- $p(x|y)$: Posterior Distribution of x given observation y
  - What we believe about x *after* we see the observation

IMAGE RESTORATION

Statistics / Estimation Recap

$$p(x|y) = \frac{p(y|x)p(x)}{\int p(y|x')p(x')dx'}$$

Maximum A Posteriori (MAP) Estimation

$$\hat{x} = \arg\max_x p(x|y)$$

- The most likely answer under the posterior distribution
- Other Estimators Possible: Try to minimize some "risk" or loss function $L(\hat{x}, x)$
  - Measures how bad an answer $\hat{x}$ is when true value is x
- Minimize Expected Risk Under Posterior

$$\hat{x} = \arg\min_x \int L(x, x')p(x'|y)dx'$$

- Let's say $L(x, x') = (x - x')^2$. What is $\hat{x}$?

$$\hat{x} = \mathbb{E}_{p(x|y)}x = \int x \ \ p(x|y)dx$$

IMAGE RESTORATION

Statistics / Estimation Recap

$$p(x|y) = \frac{p(y|x)p(x)}{\int p(y|x')p(x')dx'}$$

Maximum A Posteriori (MAP) Estimation

$$\hat{x} = \arg\max_x p(x|y)$$

How do we compute this ?

$$\hat{x} = \arg\max_x p(x|y) = \arg\max_x \frac{p(y|x)p(x)}{\int p(y|x')p(x')dx'}$$

$$= \arg\max_x p(y|x)p(x) = \arg\max_x \log[p(y|x)p(x)]$$

$$= \arg\max_x \log p(y|x) + \log p(x)$$

$$= \arg\min_x - \log p(y|x) - \log p(x)$$

(Turn this into minimization of a cost)

IMAGE RESTORATION

Statistics / Estimation Recap

Maximum A Posteriori (MAP) Estimation

$$\hat{x} = \arg\min_x \frac{(y-x)^2}{2\sigma^2} + C + \frac{(x-0.5)^2}{2} + C'$$

Back to Denoising

- $p(y|x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y-x)^2}{2\sigma^2}\right)$
- Let's choose a simple prior:
  - $p(x) = \mathcal{N}(x; 0.5, 1)$

IMAGE RESTORATION

Statistics / Estimation Recap

Maximum A Posteriori (MAP) Estimation

$$\hat{x} = \arg\min_x \frac{(y-x)^2}{2\sigma^2} + \frac{(x-0.5)^2}{2}$$

How do you compute x ?

- In this case, simpler option available. Complete the squares, express as a single quadratic $\propto (x-\hat{x})^2$.
- More generally, to minimize $C(x)$, find $C'(x) = 0$, check $C''(x) > 0$
- What is $C'(x)$?

$$C'(x) = \frac{x-y}{\sigma^2} + x - 0.5$$

IMAGE RESTORATION

Statistics / Estimation Recap

Maximum A Posteriori (MAP) Estimation

$$\hat{x} = \arg\min_{x} \frac{(y-x)^2}{2\sigma^2} + \frac{(x-0.5)^2}{2}$$

$$C'(x) = \frac{x-y}{\sigma^2} + x - 0.5 = 0$$

$$\hat{x} = \frac{y + 0.5\sigma^2}{1+\sigma^2}$$

$$C''(x) = \frac{1}{\sigma^2} + 1 > 0 \qquad \forall x$$

- Means the function is convex (quadratic functions with a positive coefficient for x^2 are convex)

IMAGE RESTORATION

Statistics / Estimation Recap

Maximum A Posteriori (MAP) Estimation

$$\hat{x} = \arg\min_{x} \frac{(y-x)^2}{2\sigma^2} + \frac{(x-0.5)^2}{2}$$

$$\hat{x} = \frac{y + 0.5\sigma^2}{1+\sigma^2}$$

- Weighted sum of observation and prior mean
- Closer to prior mean when σ^2 is high

IMAGE RESTORATION

Statistics / Estimation Recap

- Let's go beyond "single pixel images": $Y[n] = X[n] + \epsilon[n]$
- If noise is independent at each pixel

$$p(\{Y[n]\}|\{X[n]\}) = \prod_n p(Y[n]|X[n]) = \prod_n \mathcal{N}(Y[n];\ X[n], \sigma^2)$$

- Similarly, if prior $p(\{X[n]\})$ is defined to model pixels independently:

$$p(\{X[n]\}) = \prod_n \mathcal{N}(X[n];\ 0.5, 1)$$

- Product turns into sum after taking $\log$

IMAGE RESTORATION

Statistics / Estimation Recap

$$\hat{X}[n] = \arg\min_{\{X[n]\}} = \sum_n \frac{(Y[n] - X[n])^2}{2\sigma^2} + \frac{(X[n] - 0.5)^2}{2}$$

- Note: This is a minimization over multiple variables

- But since the different variables don't interact with each other, we can optimize each pixel separately

$$\hat{X}[n] = \frac{Y[n] + 0.5\sigma^2}{1 + \sigma^2} \quad \forall n$$

IMAGE RESTORATION

Multi-variate Gaussians

- Re-interpret these as multi-variate Gaussians

For $X \in \mathbb{R}^d$:

$$p(X) = \frac{1}{\sqrt{(2\pi)^d \det(\Sigma)}} \exp\left(-\frac{1}{2}(X-\mu)^T \Sigma^{-1} (X-\mu)\right)$$

- Here, $\mu \in \mathbb{R}^d$ is a mean "vector", same size as X
- Co-variance Σ is a symmetric, positive-definite matrix $d \times d$ matrix

When the different elements of X are independent, off-diagonal elements of Σ are 0.

How do we represent our prior and likelihood as a multi-variate Gaussian ?

IMAGE RESTORATION

Multi-variate Gaussians

- Represent X and Y as vectorized images

$$p(Y|X) = \mathcal{N}(Y; X, \sigma^2 I)$$

- Here the mean-vector for Y is X
- The co-variance matrix is a diagonal matrix with all diagonal entries $= \sigma^2$

IMAGE RESTORATION

Multi-variate Gaussians

$$p(Y|X) = \frac{1}{\sqrt{(2\pi)^d \det(\Sigma)}} \exp\left(-\frac{1}{2}(Y-\mu)^T \Sigma^{-1} (Y-\mu)\right)$$

- $\mu = X, \Sigma = \sigma^2 I$
- $\det(\Sigma) = (\sigma^2)^d$
- $\Sigma^{-1} = \sigma^{-2} I$

$$(Y-\mu)^T \Sigma^{-1} (Y-\mu) = (Y-X)^T \frac{1}{\sigma^2} I (Y-X) = \frac{1}{\sigma^2}(Y-X)^T \, I \, (Y-X)$$

$$= \frac{1}{\sigma^2}(Y-X)^T(Y-X) = \frac{1}{\sigma^2}\|Y-X\|^2$$

$$= \frac{1}{\sigma^2}\sum_n (Y[n]-X[n])^2$$

IMAGE RESTORATION

$$p(Y|X) = \mathcal{N}(Y; X, \sigma^2 I)$$

$$p(X) = \mathcal{N}(Y; 0.5I, I)$$

$$\hat{X} = \arg\min_X \frac{1}{\sigma}^2 \|Y - X\|^2 + \|X - 0.5\|^2$$

- But now, we can use these multi-variate distributions to model correlations and interactions between different pixels
- Let's stay with denoising, but talk about defining a prior in the Wavelet domain
- Instead of saying, individual pixel values are independent, let's say individual wavelet coefficients are independent
- Let's also put different means and variances on different wavelet coefficients
  - All the 'derivative' coefficients have zero mean
  - Variance goes up as we go to coarser levels

IMAGE RESTORATION

Let $C = W\ X$, where W represents the Wavelet transform matrix

- W is unitary, $W^{-1} = W^T$
- $X = W^T C$

Define our prior on C:

$$p(C) = \mathcal{N}(C; \mu_c, D_c)$$

- Now, D_c is a diagonal matrix, with entries equal to corresponding variances of coefficients
- Off-diagonal elements 0 implies Wavelet co-efficients are un-correlated
- A prior on C implies a prior on X
  - C is just a different representation for X

IMAGE RESTORATION

Change of Variables

- We have probability distribution on a random variable U: $p_U(U)$
- $U = f(V)$ is a one-to-one function

$$p_V(V) = p_U(f(V)) \quad ?$$

Realize, that $p(\cdot)$ are densities. So we need to account for "scaling" of the probability measure.

$$\int p_U(U)dU = \int p_U(f(V))\color{red}{dU}$$

$$dU = \det(J_f)dV$$

where $\det(J_f)_{ij} = \frac{du_i}{dv_j}$

So, $p_V(V) = p_U(f(V))\det(J_f)$

If $U = f(V) = A\ V$ is linear, $J = A$.

If A is unitary, $\det(A) = 1$.

IMAGE RESTORATION

Let $C = W\ X$, where W represents the Wavelet transform matrix

$$p(C) = \mathcal{N}(C; \mu_c, D_c)$$

$$p(X) = \frac{1}{\sqrt{(2\pi)^d \det(D_c)}} \exp\left(-\frac{1}{2}(WX - \mu_c)^T D_c (WX - \mu_c)\right)$$

$$\propto \exp\left(-\frac{1}{2}(X - W^T\mu_c)^T\ W^T D_c W\ (X - W^T\mu_c)\right)$$

$$= \mathcal{N}(X; W^T\mu_c, W^T D_c W)$$

Now, let's denoise with this prior.

IMAGE RESTORATION

$$\hat{X} = \arg\min_{X} \|X - Y\|^2 + (X - \mu)^T \Sigma^{-1} (X - \mu)$$

where, $\mu = W^T \mu_c, \Sigma = W^T D_c W$. (We're assuming noise variance is 1).

How do we minimize this ?

Take derivative, set to 0. Check second derivative is positive.

Take gradient (vector derivative), set to 0. Check that "Hessian" is positive-definite.

$y = f(X)$ is a scalar valued function of a vector $X \in \mathbb{R}^d$.

- The gradient is a vector same sized as X, with each entry being the partial derivative of y with respect to that element of X.

$$\nabla_X y = \begin{bmatrix} \frac{\partial y}{\partial X_1} \\ \frac{\partial y}{\partial X_2} \\ \frac{\partial y}{\partial X_3} \\ \vdots \end{bmatrix}$$

IMAGE RESTORATION

$y = f(X)$ is a scalar valued function of a vector $X \in \mathbb{R}^d$.

- The gradient is a vector same sized as X, with each entry being the partial derivative of y with respect to that element of X.

$$\nabla_X y = \begin{bmatrix} \frac{\partial y}{\partial X_1} \\ \frac{\partial y}{\partial X_2} \\ \frac{\partial y}{\partial X_3} \\ \vdots \end{bmatrix}$$

- The Hessian is a matrix of size $d \times d$ for $X \in \mathbb{R}^d$:

$$(H_{yx})_{ij} = \frac{\partial^2 y}{\partial X_i \partial X_j}$$

IMAGE RESTORATION

Properties of a multi-variate quadratic form

$$X^T QX - 2X^T R + S$$

where Q is a symmetric $d \times d$ matrix, R is a d-dimensional vector, S is a scalar.

- Note that this is a scalar value

$$\nabla_X = 2QX - 2R$$

- Comes from the following identities
  - $\nabla_X X^T AX = (A + A^T)X$
  - $\nabla_X X^T R = \nabla_X R^T X = R$
- The Hessian is simply given by Q (it is constant, doesn't depend on X)
- If Q is positive definite, then $\nabla_X = 0$ gives us the unique minimizer.

$\nabla_X = 0 \rightarrow 2QX - 2R = 0 \rightarrow X = Q^{-1}R$

IMAGE RESTORATION

$$\hat{X} = \arg\min_{X} \|X - Y\|^2 + (X - \mu)^T \Sigma^{-1} (X - \mu)$$

where, $\mu = W^T \mu_c, \Sigma = W^T D_c W.$

$$= \arg\min_{X} X^T (I + \Sigma^{-1}) X - 2X^T (Y + \Sigma^{-1}\mu) + \left(\|Y\|^2 + \mu^T \Sigma^{-1} \mu\right)$$

$Q = (I + \Sigma^{-1})$ is positive definite (sum of two positive-definite matrices is positive-definite).

$$\hat{X} = (I + \Sigma^{-1})^{-1} (Y + \Sigma^{-1}\mu)$$

$$\Sigma^{-1} = (W^T D_c W)^{-1} = W^{-1} D_c^{-1} W^{T^{-1}} = W^T D_c^{-1} W$$

i.e., taking inverse in the original de-correlated wavelet domain.

$$\Sigma^{-1}\mu = W^T D_c^{-1} W \mu = W^T D_c^{-1} W W^T \mu_c = W^T D_c^{-1} \mu_c$$

i.e., inverse-wavelet transform of the scaled wavelet coefficients.

IMAGE RESTORATION

$$\hat{X} = \arg\min_X X^T(I+\Sigma^{-1})X - 2X^T(Y+\Sigma^{-1}\mu) + \left(\|Y\|^2 + \mu^T\Sigma^{-1}\mu\right)$$
$$= (I+\Sigma^{-1})^{-1}(Y+\Sigma^{-1}\mu)$$

$$\Sigma^{-1} = W^T D_c^{-1} W, \quad \Sigma^{-1}\mu = W^T D_c^{-1}\mu_c$$

Note that we can't actually form $I+\Sigma^{-1}$, let alone invert it: $d \times d$ matrix, with $d =$ total number of pixels.

$$I+\Sigma^{-1} = I + W^T D_c^{-1} W$$

$$= W^T W + W^T D_c^{-1} W = W^T I W + W^T D_c^{-1} W = W^T(I+D_c^{-1})W$$

$$(I+\Sigma^{-1})^{-1} = W^T(I+D_c^{-1})^{-1}W$$

$I + D_c^{-1}$ is also diagonal, so it's inverse is just inverting diagonal elements.

$$\hat{X} = W^T(I+D_c^{-1})^{-1}W \quad (Y + W^T D_c^{-1}\mu_c)$$

IMAGE RESTORATION

$$\hat{X} = \arg\min_X X^T(I + \Sigma^{-1})X - 2X^T(Y + \Sigma^{-1}\mu) + \left(\|Y\|^2 + \mu^T\Sigma^{-1}\mu\right)$$

$$\hat{X} = W^T(I + D_c^{-1})^{-1}W \quad (Y + W^TD_c^{-1}\mu_c)$$

$$W\hat{X} = WW^T(I + D_c^{-1})^{-1}(WY + WW^TD_c{}^{-}1\mu_c)$$

Let's call $\hat{C} = W\hat{X}, C_y = WY.$

$$\hat{C} = (I + D_c^{-1})^{-1}(C_y + D_c{}^{-}1\mu_c)$$

So what we did is that we "denoised" in the wavelet domain, with the noise variance in the wavelet domain also being equal to 1.

$Y = X + \epsilon \rightarrow C_y = C + W\epsilon$

$W\epsilon$ is also a random vector with 0 mean, and covariance $= WIW^T = I.$

IMAGE RESTORATION

So far, we've been in a Bayesian setting with a prior.

$$\hat{X} = \arg\min_X -\log p(Y|X) - \log p(X)$$

- We're balancing off fidelity to observation (first term, likelihood), and what we believe about X (second term, prior)
- But we need $p(X)$ to be 'a proper' probability distribution. Sometimes, that's too restrictive.
- Instead just think of these as a "data term" (still comes from observation model), and have a generic "regularization" term.

$$\hat{X} = \arg\min_X D(X;Y) + R(X)$$

$$\hat{X} = \arg\min_X \|X - Y\|^2 + \lambda \left(\|G_x * X\|^2 + \|G_y * X\|^2\right)$$

- Here, $G_x * X$ is a vector corresponding to the image we get by convolving X with the Gaussian x-derivative filter.

IMAGE RESTORATION

$$\hat{X} = \arg\min_X \|X - Y\|^2 + \lambda \left(\|G_x * X\|^2 + \|G_y * X\|^2 \right)$$

- Data term still has the same form as for denoising.
- But now, regularization just says we want the x and y spatial derivatives of our image to be small
  - Encodes a prior that natural images are smooth.
- This isn't a prior. Because gradients aren't a "representation" of X.
- But better, because they don't enforce smoothness at alternate gradient locations.

How do we minimize this ?