Modelización e Inferencia Bayesiana aplicada a Clasificación y Procesamiento de Imágenes

Rafael Molina

Universidad de Granada

http://decsai.ugr.es/~rms

rms@decsai.ugr.es

Resumen

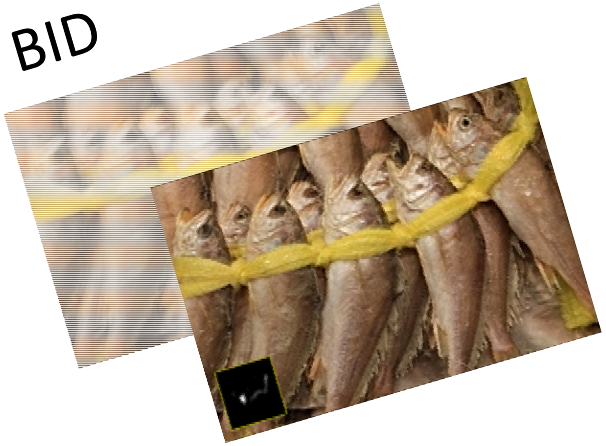
- *¿Qué comparten los siguientes problemas?*
- *Introducción (con un ejemplo antiguo)*
- *Observaciones y variables desconocidas*
- *Modelización e Inferencia Bayesianas*
- *Inferencia Bayesiana Variacional*
  - *Aproximación por Campo Medio*
  - Ejemplo: Deconvolución Ciega (BID)
  - Algoritmo EM Variational Bayesiano
  - Ejemplo: BID con estimación de parámetros
  - Conexiones con otros métodos de inferencia
  - *Acotando las distribuciones*
- *Dos ejemplos: Deconvolución + Filtrado y Clasificación*

I. ¿Qué comparten los siguientes problemas?

BID

I. ¿Qué comparten los siguientes problemas?

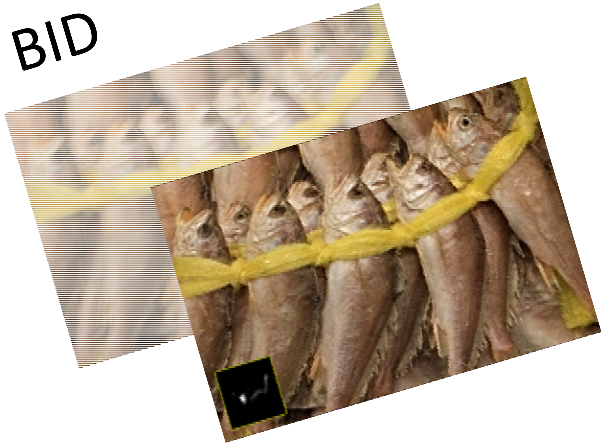
BID



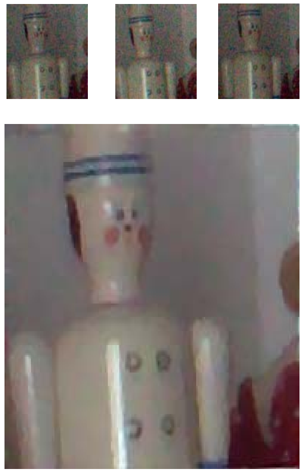
SR

I. ¿Qué comparten los siguientes problemas?

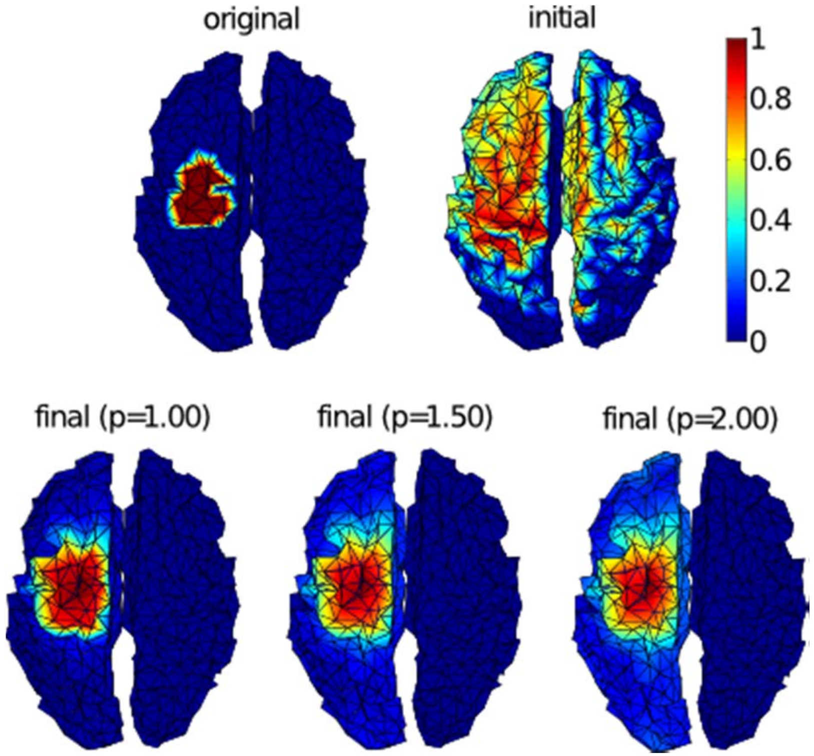
BID



SR

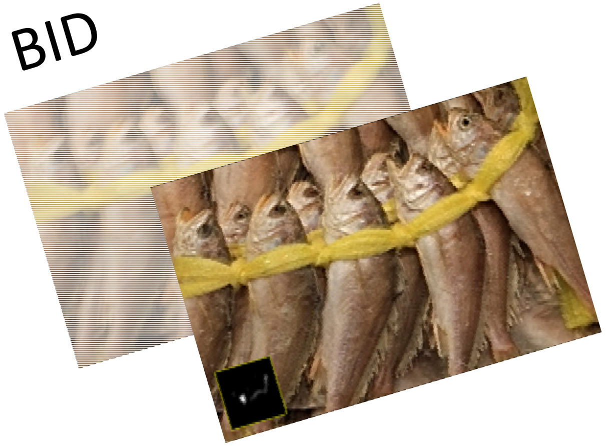


Localización de fuentes



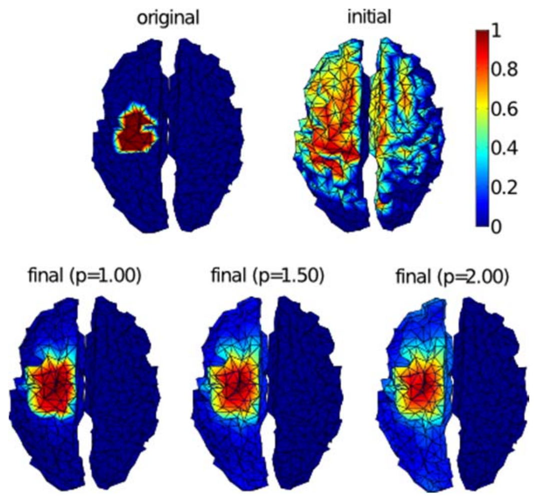
I. ¿Qué comparten los siguientes problemas?

BID

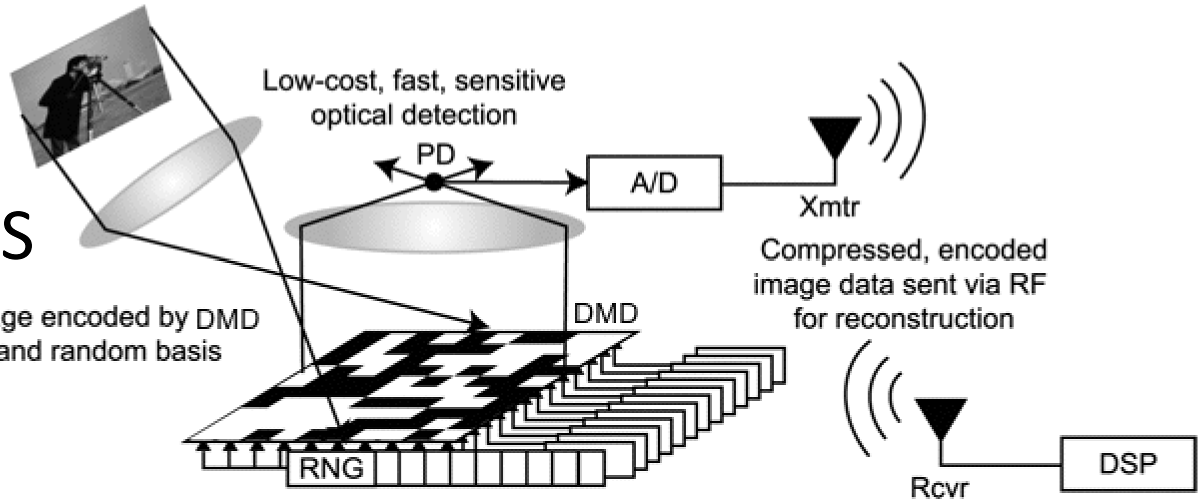


SR

Localización de fuentes

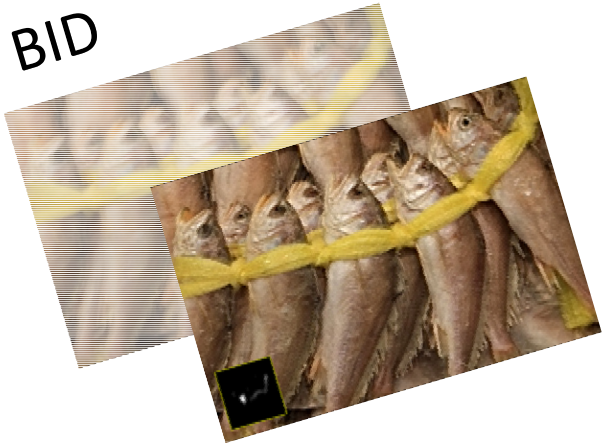


CS



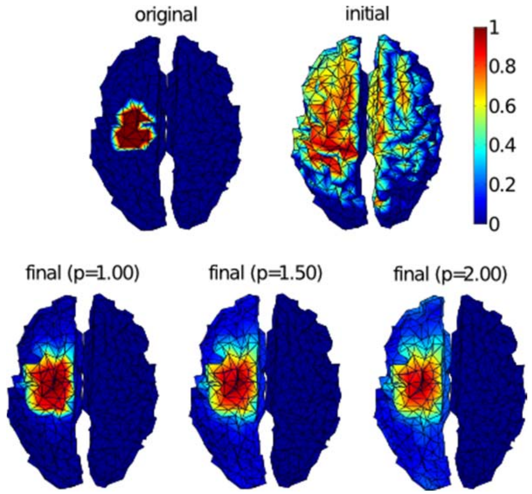
I. ¿Qué comparten los siguientes problemas?

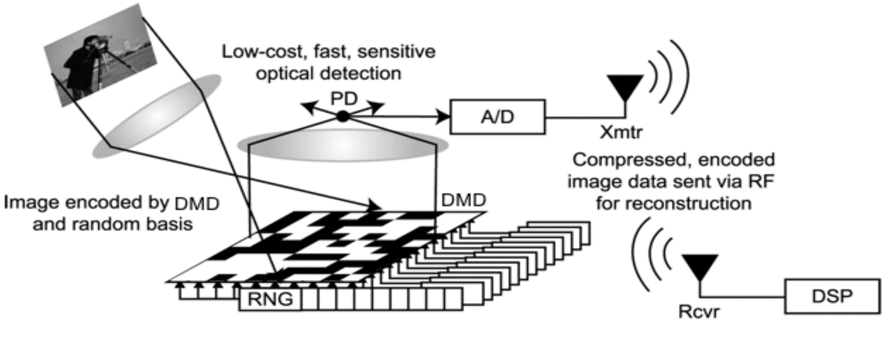
BID



SR

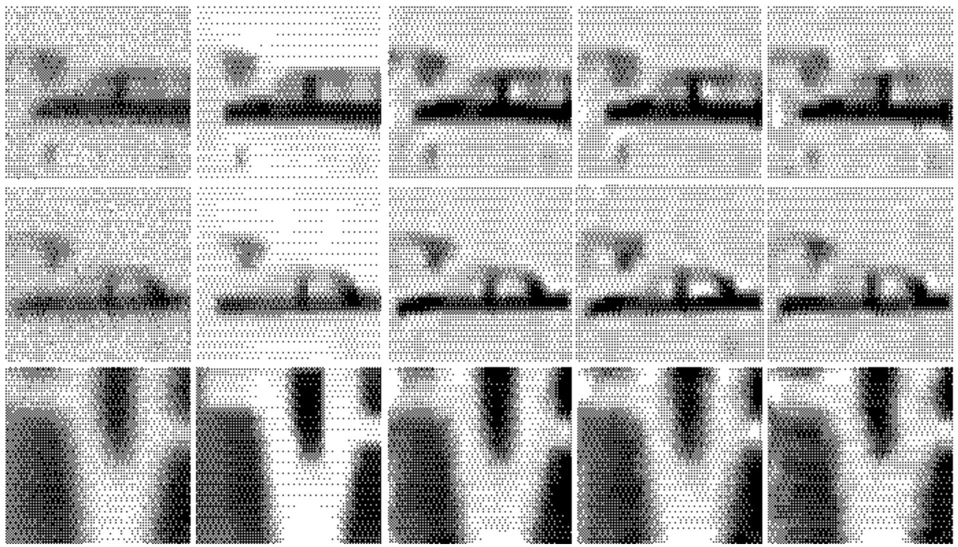
Source Location





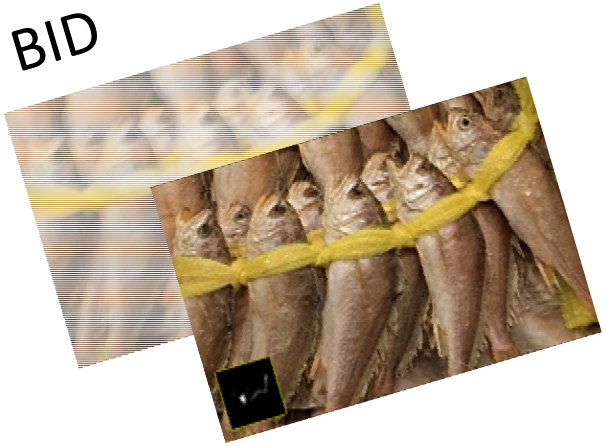
CS

PWMMI



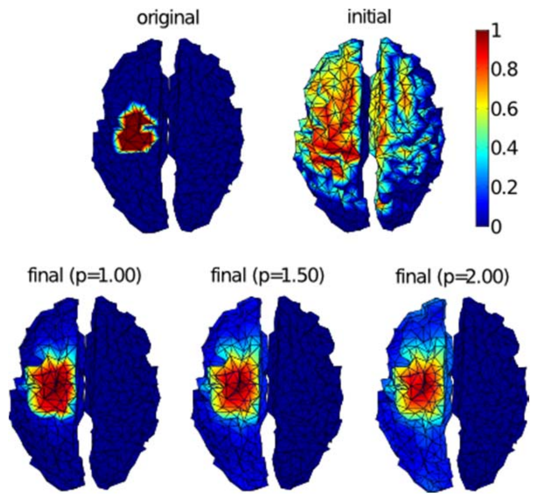
I. ¿Qué comparten los siguientes problemas?

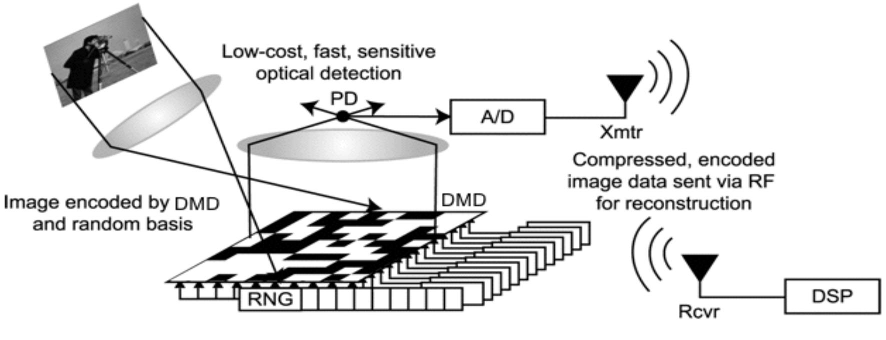
BID



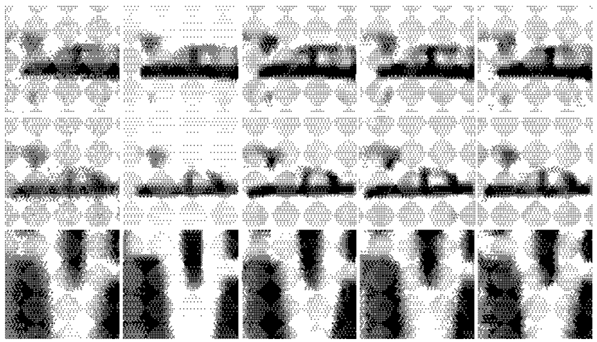
SR

Source Location





CS



PWMMI

CP

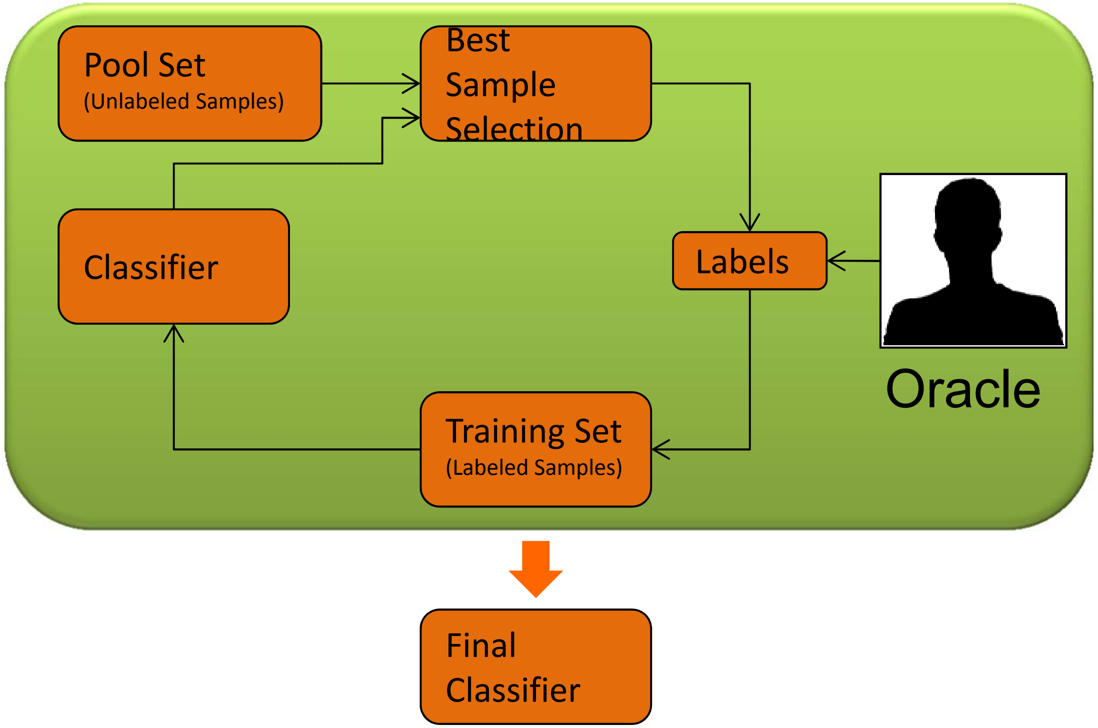
I. ¿Qué comparten los siguientes problemas?

RPCA

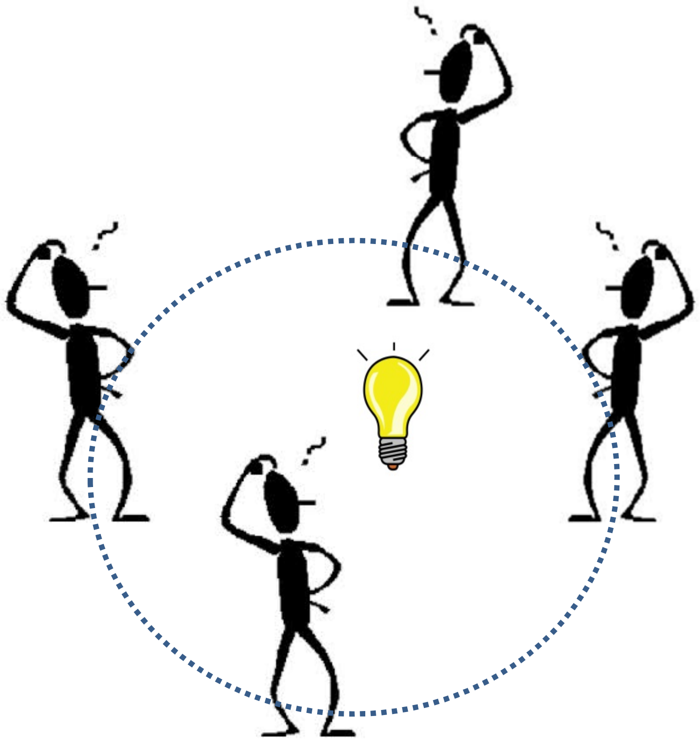
GRPCA
Detección de Anomalías

I. ¿Qué comparten los siguientes problemas?

Aprendizaje Activo



Crowdsourcing



I. ¿Qué comparten los siguientes problemas?

La solución a estos problemas (y a muchos otros) puede abordarse utilizando ….

Modelización e Inferencia Bayesiana

II. *Introducción (con un ejemplo antiguo)*

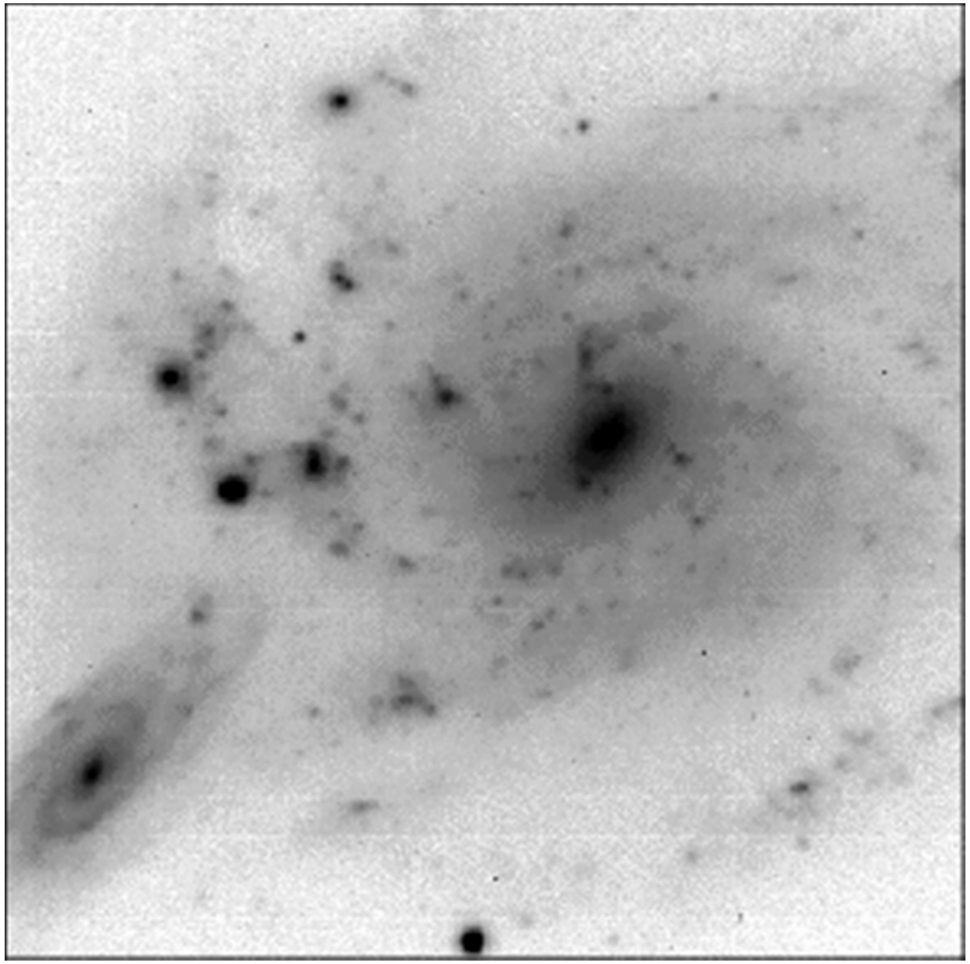


Imagen proporcionada por el Instituto de Astrofísica de Andalucía (IAA) en 1987

Objetivo: Quitar el emborronamiento (deconvolucionar) la imagen

¿Cómo podemos abordar el problema?

$$\mathbf{w} = \ln(\mathbf{x} + \mathrm{const})$$

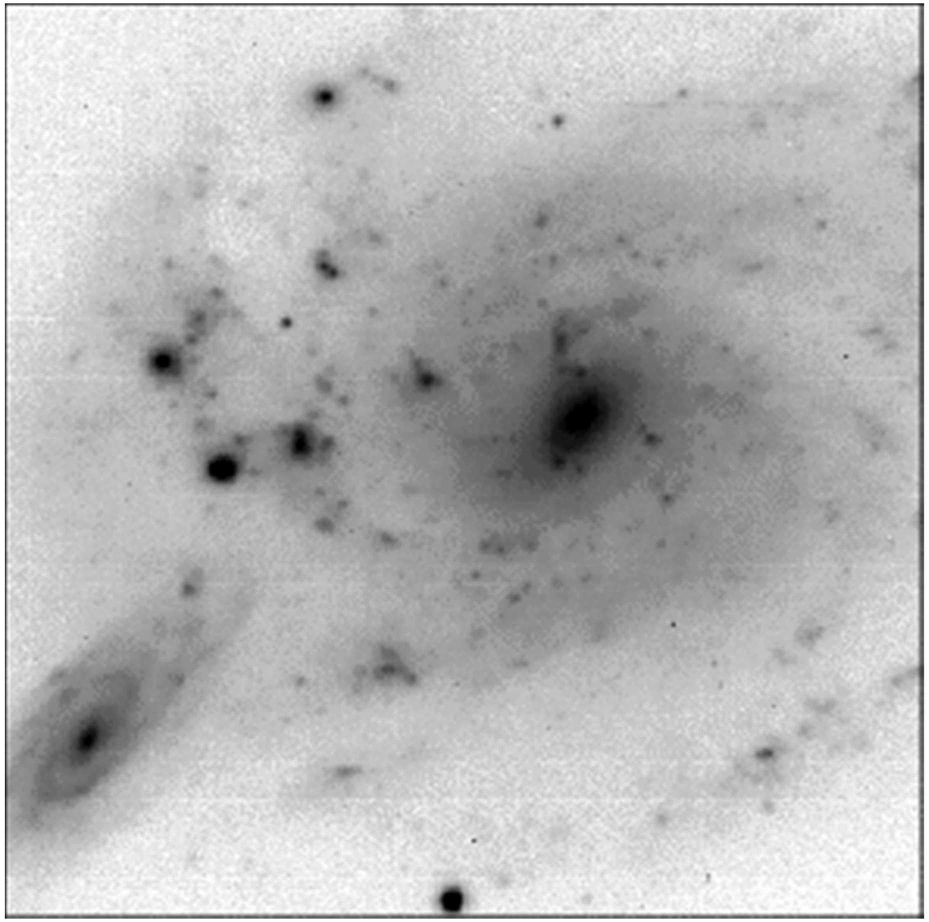
Utilizamos como distribuciones a priori sobre la imagen transformada modelos condicionales o simultáneos autoregresivos (CAR o SAR)

Ruido gaussiano dependiente de la señal

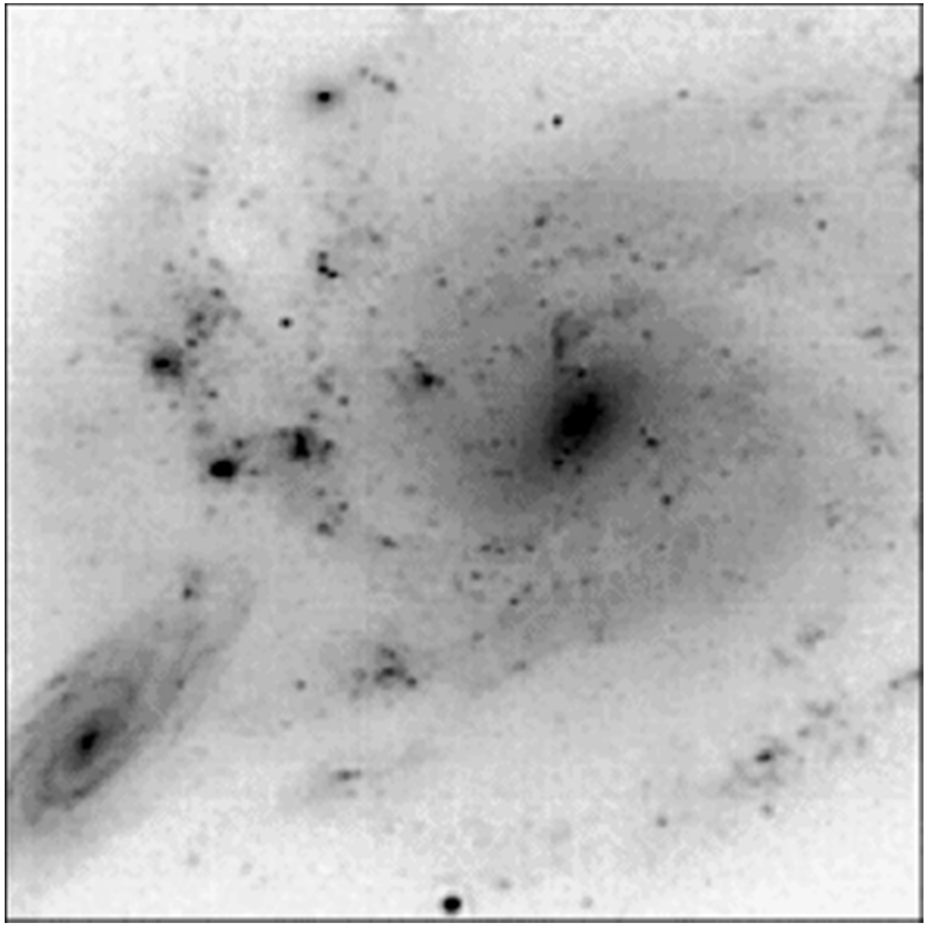
Inferencia (no mucho más que MAP en aquel tiempo)

$$\hat{\mathbf{w}} = \arg\max_{\mathbf{w}} p(\mathbf{w}|\mathbf{y}) = \arg\max_{\mathbf{w}} \mathrm{p}(\mathbf{w})\mathrm{p}(\mathbf{y}|\mathbf{w})$$

- R. Molina, and B.D. Ripley, “Using Spatial Models as Priors in Astronomical Image Analysis”, Journal of Applied Statistics, vol. 16, 193-206, 1989.
- R. Molina, A. del Olmo, J. Perea, and B.D. Ripley, “Bayesian Deconvolution in Optical Astronomy”, The Astronomical Journal, vol. 103, 666-675, 1992.
- R. Molina, “On the Hierarchical Bayesian Approach to Image Restoration. Applications to Astronomical Images”, IEEE PAMI. vol. 16, 1122-1128, 1994.



observación



restauración

III. *Observaciones y variables desconocidas*

Formulación inicial muy simplificada

Variables observadas

$$\mathbf{y}$$

Variables desconocidas (latentes)

$$\mathbf{z}$$

Normalmente z notará el conjunto de variables latentes y no contendrá los parámetros del modelo

En esta formulación simplificada suponemos que conocemos el emborronamiento y los parámetros del modelo: tanto del modelo a priori como del de observación

Ejemplo en restauración de imágenes

observada

$$\mathbf{y} = \mathbf{H}\mathbf{z} + \epsilon$$

conocido

a estimar

IV. Modelización e Inferencia Bayesianas

Modelización Bayesiana

$$\mathrm{p}(\mathbf{z}, \mathbf{y}) = \mathrm{p}(\mathbf{z})\mathrm{p}(\mathbf{y}|\mathbf{z})$$

Inferencia Bayesiana

Basada en encontrar, estimar,....

$$\mathrm{p}(\mathbf{z}|\mathbf{y})$$

Es mucho más que el MAP

$$\hat{\mathbf{z}} = \arg\max_{\mathbf{z}} \mathrm{p}(\mathbf{z})\mathrm{p}(\mathbf{y}|\mathbf{z})$$

IV. Modelización e Inferencia Bayesianas

¿Podemos encontrar siempre

$$\mathrm{p}(\mathbf{z}|\mathbf{y}) \quad ?$$

No!!.

Veamos como aproximar esta distribución

V. Inferencia Variacional

Consideremos dos distribuciones de probabilidad $\mathrm{q}(\mathrm{u})$ and $\mathrm{p}(\mathrm{u})$ y analicemos su divergencia de Kullback-Leibler

$$\boxed{\mathrm{KL}(\mathrm{q}(u) \,\|\, \mathrm{p}(u)) = \int \mathrm{q}(u) \ln \frac{\mathrm{q}(u)}{\mathrm{p}(u)} du}$$

Ya que para r > 0, $\ln r \leq r - 1$

$$\mathrm{KL}(\mathrm{q}(u) \,\|\, \mathrm{p}(u)) \;=\; -\int \mathrm{q}(u) \ln \frac{\mathrm{p}(u)}{\mathrm{q}(u)} du \geq 0$$

$$\mathrm{KL}(\mathrm{q}(u) \,\|\, \mathrm{p}(u)) \;=\; 0$$

Si ambas distribuciones coinciden casi seguramente

IV. Inferencia Variacional

¿Qué deberíamos minimizar para encontrar $\mathrm{q}(\mathbf{z})$,

$$\mathrm{KL}(\mathrm{q}(\mathbf{z})||\mathrm{p}(\mathbf{z}|\mathbf{y}))$$

o

$$\mathrm{KL}(\mathrm{p}(\mathbf{z}|\mathbf{y})||\mathrm{q}(\mathbf{z}))$$

?

IV. Inferencia Variacional

La KL inversa, también llamada I-projection o proyección de información, intenta ajustar bien donde está la masa de la distribución.

$$\mathrm{KL}(\mathrm{q}(\mathbf{z})||\mathrm{p}(\mathbf{z}|\mathbf{y})) = \int \mathrm{q}(\mathbf{z}) \ln \left(\frac{\mathrm{q}(\mathbf{z})}{\mathrm{p}(\mathbf{z}|\mathbf{y})} \right) \mathrm{d}\mathbf{z}$$

- donde $\mathrm{p}(\mathbf{z}|\mathbf{y})$ es pequeña, una buena aproximación, $\mathrm{q}(\mathbf{z})$, también será pequeña
- La KL inversa fuerza el cero en q(z) (if p(z|y)=0 entonces q(z)=0). Normalmente infra-estimará el soporte de p(z|y)

IV. Inferencia Variacional

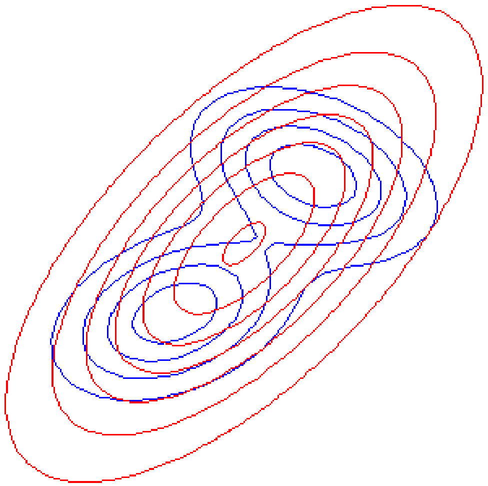
La KL hacia delante, también llamada M-projection o proyección de momentos, se define mediante.

$$\mathrm{KL}(\mathrm{p}(\mathbf{z}|\mathbf{y})||\mathrm{q}(\mathbf{z})) = \int \mathrm{p}(\mathbf{z}|\mathbf{y}) \ln\left(\frac{\mathrm{p}(\mathbf{z}|\mathbf{y})}{\mathrm{q}(\mathbf{z})}\right) \mathrm{d}\mathbf{z}$$

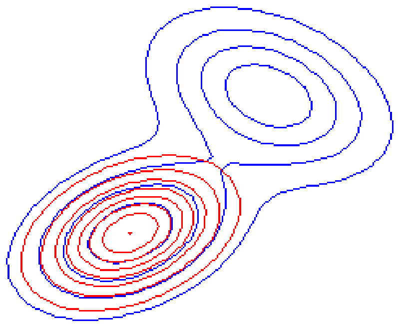
- Si q(z)=0 y p(z|y)>0 la divergencia es infinita. La KL hacia delante evita ceros en q(z). Normalmente sobre-estima el soporte de p(z|y).
- La KL hacia delante se utiliza en Propagación de la Esperanza (EP), más un poco más tarde.

IV. Inferencia Variacional

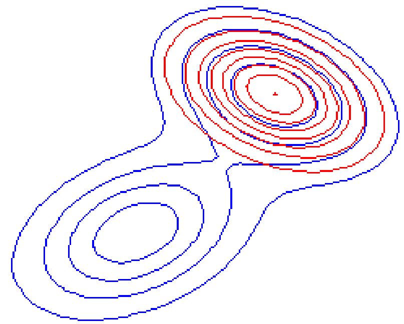
KL hacia delante o atrás



En azul, contornos de la distribución original. En rojo aproximación unimodal hacia delante



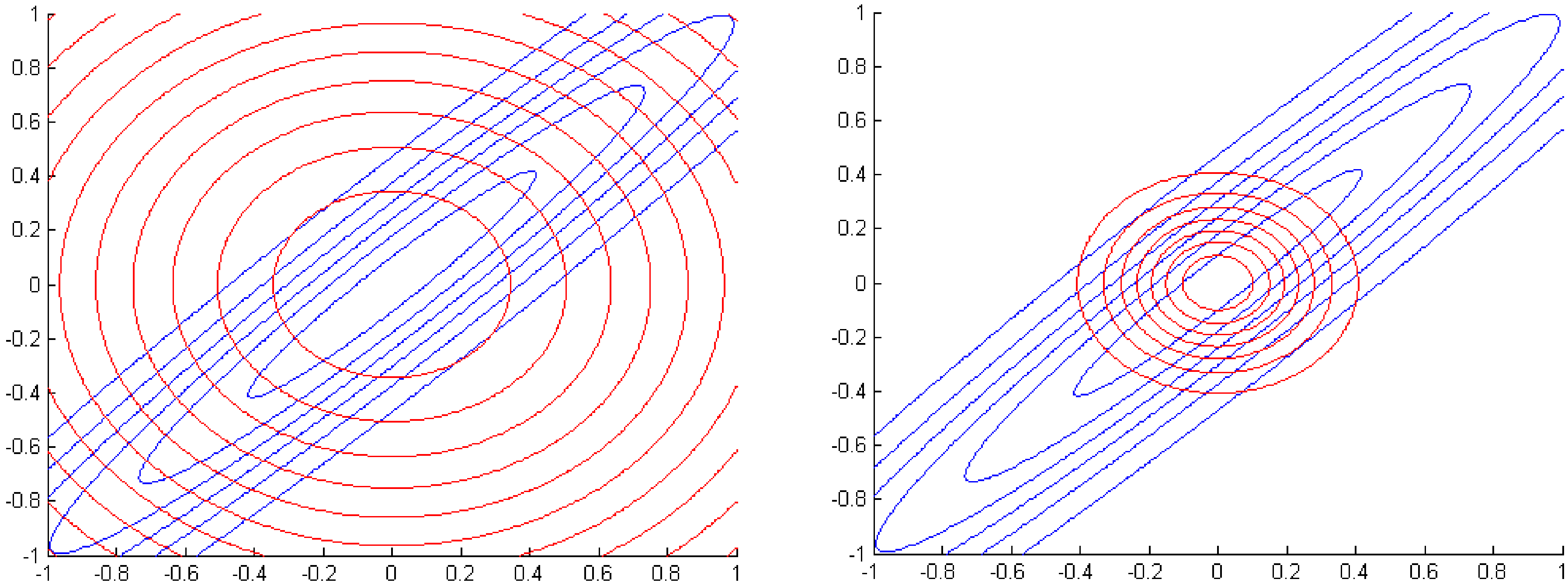
En azul, contornos de la distribución original. En rojo, aproximación unimodal hacia atrás



En azul, contornos de la distribución original. En rojo, aproximación unimodal hacia atrás

IV. Inferencia Variacional

KL hacia delante o atrás



En azul, contornos de la distribución original. En rojo aproximación unimodal hacia delante

En azul, contornos de la distribución original. En rojo, aproximación unimodal hacia atrás

IV. Inferencia Variacional

Problema: No conocemos la distribución a posteriori que aparece en el denominador.

De

$$\mathrm{KL}(\mathrm{q}(\mathbf{z})||\mathrm{p}(\mathbf{z}|\mathbf{y})) = \int \mathrm{q}(\mathbf{z}) \ln \left(\frac{\mathrm{q}(\mathbf{z})}{\mathrm{p}(\mathbf{z}|\mathbf{y})} \right) \mathrm{d}\mathbf{z}$$

tenemos

Funcional para ser minimizado

$$\mathrm{KL}(\mathrm{q}(\mathbf{z})||\mathrm{p}(\mathbf{z}|\mathbf{y})) = \int \mathrm{q}(\mathbf{z}) \ln \left(\frac{\mathrm{q}(\mathbf{z})}{\mathrm{p}(\mathbf{z},\mathbf{y})} \right) \mathrm{d}\mathbf{z} + \ln \mathrm{p}(\mathbf{y})$$

Así pues, para encontrar q(z) no necesitamos la verdadera distribución a posteriori, sólo la distribución conjunta

IV. Inferencia Variacional

Observa también que de

$$0 \leq \mathrm{KL}(\mathrm{q}(\mathbf{z})||\mathrm{p}(\mathbf{z}|\mathbf{y})) = \int \mathrm{q}(\mathbf{z}) \ln \left(\frac{\mathrm{q}(\mathbf{z})}{\mathrm{p}(\mathbf{z}, \mathbf{y})} \right) \mathrm{d}\mathbf{z} + \ln \mathrm{p}(\mathbf{y})$$

$$\ln \mathrm{p}(\mathbf{y}) \geq - \int \mathrm{q}(\mathbf{z}) \ln \left(\frac{\mathrm{q}(\mathbf{z})}{\mathrm{p}(\mathbf{z}, \mathbf{y})} \right) \mathrm{d}\mathbf{z}$$

Tenemos una cota inferior de ln p(y).
Puede usarse para estimar los parámetros del modelo

IV. Inferencia Variacional

Pregunta: ¿Cómo seleccionamos la aproximación?.

- Una posible solución es imponer una forma específica paramétrica a $\mathrm{q}(\mathbf{z})$, por ejemplo ser Gaussiana

- Otra hipótesis, usada frecuentemente, es

$$\mathrm{q}(\mathbf{z}) = \prod_{i=1} \mathrm{q}_i(\mathbf{z}^i),$$

Conjuntos de variables disjuntas en $\mathbf{z}$,

Esta forma factorizada de inferencia variacional se llama en Física Teoría del Campo Medio

IV. Inferencia Variacional

Campo Medio

La solución óptima para cada factor es

$$q_i(\mathbf{z}^i) \propto \exp\left\{\mathrm{E}_{j\neq i}\left[\ln \mathrm{p}(\mathbf{z},\mathbf{y})\right]\right\}$$

- Ésta es probablemente la ecuación VB más importante.
- Necesitamos calcular o aproximar la media

Media sobre todas las variables excluyendo la distribución sobre $\mathbf{z}^i$

La ecuación anterior lleva al siguiente procedimiento de optimización: la distribución de cada factor se actualiza usando las distribuciones más reciente de todos los otros factores.

Ejemplo (simple) de BID

$$\mathbf{z} = (\mathbf{x}, \mathbf{h})$$

Modelo de imagen

$$\mathrm{p}(\mathbf{x}) \propto \exp\left[-\frac{\alpha_1}{2}\mathbf{x}^t\mathbf{C}_1^t\mathbf{C}_1\mathbf{x}\right]$$

Modelo de emborronamiento

$$\mathrm{p}(\mathbf{h}) \propto \exp\left[-\frac{\alpha_2}{2}\mathbf{h}^t\mathbf{C}_2^t\mathbf{C}_2\mathbf{h}\right]$$

- Restricciones de suavidad
- Penalizan imágenes y emborronamientos no suaves

Modelo de observación

$$\mathrm{p}(\mathbf{y}|\mathbf{x}, \mathbf{h}) \propto \exp\left[-\frac{\beta}{2} \parallel \mathbf{y} - \mathbf{H}\mathbf{x} \parallel^2\right]$$

$\alpha_1, \alpha_2, \beta$ Se suponen conocidos

Objetivo: estimar la imagen y el emborronamiento

Ejemplo (simple) de BID

Nos gustaría encontrar

$$\mathrm{p}(\mathbf{x},\mathbf{h},|\mathbf{y}) = \frac{\mathrm{p}(\mathbf{y},\mathbf{x},\mathbf{h})}{\mathrm{p}(\mathbf{y})}$$

Sin embargo, no podemos calcular

$$\mathrm{p}(\mathbf{y}) = \int\int p(\mathbf{y},\mathbf{x},\mathbf{h})\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{h}$$

Buscamos la aproximación por campo medio

$$\mathrm{p}(\mathbf{x},\mathbf{h}|\mathbf{y}) \approx \mathrm{q}(\mathbf{x})\mathrm{q}(\mathbf{h})$$

Ejemplo (simple) de BID

Por la teoría VB sabemos

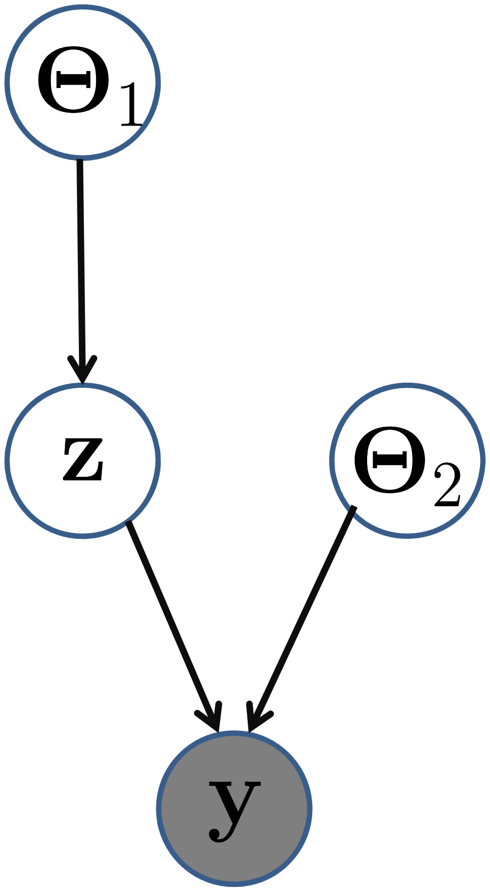
$$
\begin{aligned}
\ln \mathrm{q}(\mathbf{x}) &= \mathbf{E}_{\mathrm{q}(\mathbf{h})}\left[\ln \mathrm{p}(\mathbf{y}, \mathbf{h}, \mathbf{x})\right] + \mathrm{const} \\
&= -\frac{\alpha_1}{2} \parallel \mathbf{C}_1 \mathbf{x} \parallel^2 - \frac{\beta}{2} \mathrm{E}[\parallel \mathbf{y} - \mathbf{H}\mathbf{x} \parallel^2]_{\mathrm{q}^k(\mathbf{h})} \\
&+ \mathrm{const}
\end{aligned}
$$

- La aproximación de la distribución a posteriori de la imagen es una distribución gaussiana
- Lo mismo ocurre con el emborronamiento

IV. Inferencia Variacional

Algoritmo EM variacional Bayesiano

$$\mathrm{p}(\boldsymbol{\Theta}, \mathbf{y}) = \mathrm{p}(\boldsymbol{\Theta}_1)\mathrm{p}(\boldsymbol{\Theta}_2)\mathrm{p}(\mathbf{z}|\boldsymbol{\Theta}_1)\mathrm{p}(\mathbf{y}|\mathbf{z}, \boldsymbol{\Theta}_2)$$



$\boldsymbol{\Theta} = (\boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2)$ Parámetros del modelo

$\mathbf{z}$ Variables latentes

Frecuentemente podemos utilizar estimadores puntuales para $\boldsymbol{\Theta}$ porque su distribución a posteriori tiene poca incertidumbre

IV. Inferencia Variacional

EM

Paso-E: Para $\mathbf{z}$ calcula $\quad \mathrm{p}(\mathbf{z}|\mathbf{y}, \boldsymbol{\Theta}^k)$

Paso-M: Maximiza en $\boldsymbol{\Theta}$ $\quad \mathbf{E}[\ln \mathrm{p}(\mathbf{y}, \mathbf{z}, \boldsymbol{\Theta})]_{\mathrm{p}(\mathbf{z}|\mathbf{y}, \boldsymbol{\Theta}^k)}$

VBEM

$$\mathrm{p}(\mathbf{z}, \boldsymbol{\Theta}|\mathbf{y}) \approx \mathrm{q}(\boldsymbol{\Theta})\mathrm{q}(\mathbf{z})$$

Paso-E: Calcula $\ln \mathrm{q}^k(\mathbf{z}) \propto \mathbf{E}[\ln \mathrm{p}(\mathbf{y}, \mathbf{z}, \boldsymbol{\Theta})]_{\mathrm{q}^k(\boldsymbol{\Theta})}$

Paso-M: Calcula $\ln \mathrm{q}^{k+1}(\boldsymbol{\Theta}) \propto \mathbf{E}[\ln \mathrm{p}(\mathbf{y}, \mathbf{z}, \boldsymbol{\Theta})]_{\mathrm{q}^k(\mathbf{z})}$

Ejemplo BID con estimación de parámetros

Modelo de imagen $$\mathrm{p}(\mathbf{x}|\alpha_1) \propto \alpha_1^{N/2} \exp\left[-\frac{\alpha_1}{2}\mathbf{x}^t\mathbf{C}_1^t\mathbf{C}_1\mathbf{x}\right]$$

Modelo de emborronamiento $$\mathrm{p}(\mathbf{h}|\alpha_2) \propto \alpha_2^{M/2} \exp\left[-\frac{\alpha_2}{2}\mathbf{h}^t\mathbf{C}_2^t\mathbf{C}_2\mathbf{h}\right]$$

Modelo de observación $$\mathrm{p}(\mathbf{y}|\mathbf{x},\mathbf{h},\beta) \propto \beta^{N/2} \exp\left[-\frac{\beta}{2} \parallel \mathbf{y} - \mathbf{H}\mathbf{x} \parallel^2\right]$$

Modelo para los parámetros

$$\theta \in (\alpha_1, \alpha_2, \beta)$$

$$\mathrm{p}(\theta) = \Gamma(\theta|a_\theta^o, b_\theta^o) = \frac{(b_\theta^o)^{a_\theta^o}}{\Gamma(a_\theta^o)}\theta^{a_\theta^o - 1}\exp[-b_\theta^o\,\theta],$$

Objetivo: estimación de la imagen, emborronamiento y parámetros

Ejemplo BID con estimación de parámetros

Utilizamos VBEM

$$\mathrm{p}(\mathbf{\Theta}, \mathbf{x}, \mathbf{h}|\mathbf{y}) \approx \mathrm{q}(\mathbf{\Theta})\mathrm{q}(\mathbf{x})\mathrm{q}(\mathbf{h})$$

Puede probarse fácilmente que

- La aproximación de la distribución a posteriori de la imagen es gaussiana
- Lo mismo es cierto para la distribución a posteriori del emborronamiento
- La aproximación de la distribución a posteriori de cada parámetro es una distribución Gamma

IV. Inferencia Variacional

Relaciones con otros métodos de inferencia

- Si las distribuciones $q(\mathbf{z})$ y $q(\boldsymbol{\Theta})$ son degeneradas (toman un valor con probabilidad uno) obtenemos el **estimador MAP.**

- Si la distribución $q(\boldsymbol{\Theta})$ es degenerada y podemos calcular $p(\mathbf{z}|\boldsymbol{\Theta}, \mathbf{y})$ tenemos el **algoritmo EM**

IV. Inferencia Variacional

Relaciones con otros métodos de inferencia

- VB también está relacionada con Loopy Belief propagation (LBP)

- ¿Podemos utilizar

$$\mathrm{KL}(\mathrm{p}(\boldsymbol{\Theta}|\mathbf{y}) \parallel (\mathrm{q}(\boldsymbol{\Theta})) = \int \mathrm{p}(\boldsymbol{\Theta}|\mathbf{y}) \ln \left(\frac{\mathrm{p}(\boldsymbol{\Theta}|\mathbf{y})}{\mathrm{q}(\boldsymbol{\Theta})} \right) \mathrm{d}\boldsymbol{\Theta} \quad ?$$

Sí pero con cuidado: Algoritmo **Expectation Propagation algorithm (EP)**

Ver

Z. Chen, S. D. Babacan, R. Molina, and A. K. Katsaggelos, "Variational Bayesian Methods For Multimedia Problems", IEEE Transactions on Multimedia, 1000-1017, June, 2014.

V. Acotando las distribuciones.

Empecemos con un ejemplo

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \boldsymbol{\epsilon} \qquad \boldsymbol{\epsilon} \sim \mathcal{N}(\underline{0}, \beta^{-1}\mathbf{I})$$

$$\mathrm{p}(\mathbf{x}|\alpha) \propto \frac{1}{Z_{\mathrm{TV}}(\alpha)} \exp\left[-\alpha \mathrm{TV}(\mathbf{x})\right]$$

$$\mathrm{TV}(\mathbf{x}) = \sum_i \sqrt{(\Delta_i^h(\mathbf{x}))^2 + (\Delta_i^v(\mathbf{x}))^2}$$

$\alpha, \beta, \mathbf{H}$ Conocidos por simplicidad

- Nuestro objetivo es estimar la distribución a posteriori de $\mathbf{x}$. Desgraciadamente no podemos manejar fácilmente

$$\ln \mathrm{q}(\mathbf{x}) = -\frac{\beta}{2} \parallel \mathbf{y} - \mathbf{H}\mathbf{x} \parallel^2 - \alpha TV(\mathbf{x}) + \mathrm{const}$$

V. Acotando las distribuciones.

Podemos usar

$$\mathrm{TV}(\mathbf{x}) \leq \sum_i \frac{(\Delta_i^h(\mathbf{x}))^2 + (\Delta_i^v(\mathbf{x}))^2 + u_i}{2\sqrt{u_i}}, \quad u_i > 0$$

y definir

$$M(\mathbf{x}, \mathbf{u}) = \frac{1}{Z(\alpha)} \exp[-\frac{\alpha}{2} \sum_i \frac{(\Delta_i^h(\mathbf{x}))^2 + (\Delta_i^v(\mathbf{x}))^2 + u_i}{\sqrt{u_i}}] \frac{1}{Z(\beta)} \exp[-\frac{\beta}{2} \parallel \mathbf{y} - \mathbf{H}\mathbf{x} \parallel^2]$$

Entonces

$$\int \mathrm{q}(\mathbf{x}) \ln \frac{\mathrm{q}(\mathbf{x})}{\frac{1}{Z(\alpha)} \exp[-\alpha TV(\mathbf{x})] \frac{1}{Z(\beta)} \exp[-\frac{\beta}{2} \parallel \mathbf{y} - \mathbf{H}\mathbf{x} \parallel^2]} \mathrm{d}\mathbf{x}$$

$$\leq \inf_{\mathbf{u} > \mathbf{0}} \int \mathrm{q}(\mathbf{x}) \ln \frac{\mathrm{q}(\mathbf{x})}{M(\mathbf{x}, \mathbf{u})} \mathrm{d}\mathbf{x}$$

Puede realizarse la minimización en u y q(x)

V. Acotando las distribuciones

Idea general

- Si en

$$q_i(\mathbf{z}^i) \propto \exp\left\{\mathrm{E}_{j\neq i}\left[\ln p(\mathbf{z}, \mathbf{y})\right]\right\}$$

No podemos calcular la esperanza, utilizaremos cotas sobre las distribuciones “difíciles”.

V. Acotando las distribuciones.

Una cota para el modelo de observación

El sigmoide logístico aparece frecuentemente en problemas de clasificación

$$\sigma(a) = \frac{1}{1 + \exp(-a)}$$

Una cota cuadrática que conduce a una cota superior manejable de la divergencia

$$\sigma(a) \geq \sigma(b) \exp\left((a-b)/2 - \lambda(b)(a^2 - b^2)\right),$$

a, b números reales y

$$\lambda(b) = \frac{1}{2b}\left(\sigma(b) - \frac{1}{2}\right)$$

Puede extenderse a K clases

V. Acotando las distribuciones.

Representación Variacional de distribuciones a priori ralas (sparse)

Distribuciones Super-Gaussianas (SG)

Consideremos distribuciones unidimensionales de la forma

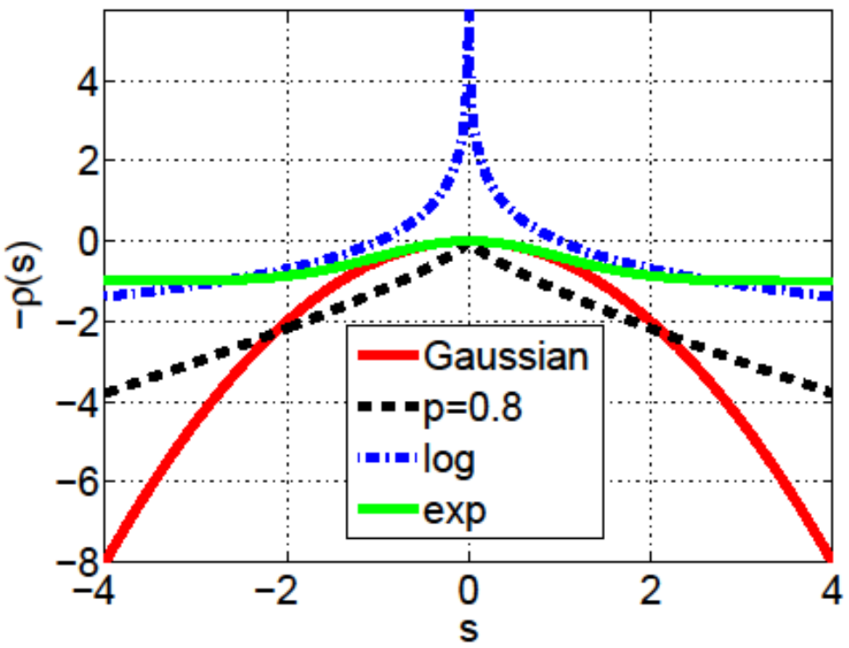
$$\mathrm{p}(u) = \frac{1}{Z}\exp\left(-\rho\left(u\right)\right)$$

- Si $\rho(\sqrt{u})$ es creciente y cóncava en $\mathcal{R}^{+}$, es decir, el cociente $\rho(u)/u$ es decreciente en $\mathcal{R}^{+}$ decimos que la distribución p(u) es SG.
- Las distribuciones SG promueven soluciones sparse

V. Acotando las distribuciones.

Ejemplos de distribuciones super-gaussianas

$\rho(s)$	$\rho'(s)/\lvert s\rvert$
$\frac{1}{2}\lvert s\rvert^2$	1
$\frac{1}{p}\lvert s\rvert^p$	$\lvert s\rvert^{p-2}$
$\log\lvert s\rvert$	$\lvert s\rvert^{-2}$
$-\sigma_r \exp\left(-\frac{\lvert s\rvert^2}{2\sigma_r}\right)$	$\exp\left(-\frac{\lvert s\rvert^2}{2\sigma_r}\right)$



V. Acotando las distribuciones

Las distribuciones SG admiten una representación muy interesante

$$\rho(u) = \inf_{\xi > 0} \frac{1}{2}\xi\, u^2 - \rho^*\left(\frac{1}{2}\xi\right)$$

$$\Rightarrow \rho(u) \leq \frac{1}{2}\xi\, u^2 - \rho^*\left(\frac{1}{2}\xi\right)$$

Función conjugada

donde $$\rho^*\left(\frac{1}{2}\xi\right) = \inf_{u} \frac{1}{2}\xi\, u^2 - \rho(u)$$

- Dado u_0, el anterior ínfimo se alcanza en $\xi_0 = \frac{\rho'(u_0)}{u_0}$. No necesitamos, por tanto, conocer ρ^*
- El mismo ξ puede ser utilizado para un conjunto de us

V. Acotando las distribuciones

Representación variacional de distribuciones sparse

<u>Mezcla de Gaussianas Escaladas (SMG)</u>

$\mathcal{N}(0, \xi^{-1})$

$$\mathrm{p}(u) = \int \mathrm{p}(u|\xi)\, \mathrm{p}(\xi)\, \mathrm{d}\xi$$

La representación SMG requiere monotonicidad completa de $\mathrm{p}(\sqrt{u})$

Una función $f(s)$ es completamente monótona si sus derivadas satisfacen $(-1)^n f^{(n)}(s) \geq 0$ para todo n=0,1,2,....

La representación SMG es un poco más restrictiva que la SG.

V. Acotando las distribuciones.

Ejemplos de distribuciones SMG

$$\mathcal{N}(0,\xi^{-1}) \quad \Gamma(a,b)$$

$$\mathrm{p}(u) = \int \mathrm{p}(u|\xi)\,\mathrm{p}(\xi)\,\mathrm{d}\xi \propto \left(a + \frac{u^2}{2}\right)^{-(b+\frac{1}{2})}$$

Para a=0, b=0 obtenemos la distribución a priori log

$$\mathrm{p}(u) = \int \mathrm{p}(u|\xi)\,\mathrm{p}(\xi)\,\mathrm{d}\xi = \frac{\lambda^{1/2}}{2}\exp\left[-\sqrt{\lambda}|u|\right]$$

Para $\mathcal{N}(0,\xi^{-1})$ y $\mathrm{I}\Gamma(1,\lambda/2)$ obtenemos la distribución de Laplace

V. Acotando las distribuciones

- ¿Necesitamos conocer $\mathrm{p}(\xi)$ en la representación SMG? No, ya que sólo necesitamos su media

$$<\xi> = \frac{\rho'(\nu)}{\nu} \text{ for } \nu = \sqrt{<u^2>}$$

Para relaciones entre modelos SG y SMG y como construir $\rho(\cdot)$ ver

- S.D. Babacan, R. Molina, M.N. Do, y A.K. Katsaggelos, "Blind deconvolution with general sparse image priors", *ECCV*, 341-355, 2012.

Ver también

- M. Vega, R. Molina, and A.K. Katsaggelos, "Parameter Estimation in Bayesian Blind Deconvolution with Super Gaussian Image Priors", EUSIPCO 2014,
- X. Zhou, R. Molina, F. Zhou, and A.K. Katsaggelos, "Fast iteratively reweighted least squares for Lp regularized image deconvolution and reconstruction", ICIP 2014.

VI. Dos ejemplos

Deconvolución ciega

A partir de una observación borrosa, estimar el emborronamiento y la imagen original

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \boldsymbol{\epsilon}$$

Aplicamos K filtros LSI basados en diferencias y obtenemos

$$\underbrace{\mathrm{L_k}\mathbf{y}}_{\mathbf{y}_k} = \mathbf{H}\underbrace{\mathrm{L_k}\mathbf{x}}_{\mathbf{z}_k} + \boldsymbol{\epsilon}_k, \quad k = 1, .., K$$

- Usamos SMG sobre $\mathbf{z}_k$, y una distribución degenerada sobre $\mathbf{H}$ and utilizamos VB para estimar el emborronamiento.
- Una vez que $\mathbf{H}$ ha sido estimada, usamos $\mathbf{y}$ and las distribuciones sobre $\mathrm{L_k}\mathbf{x}$ para estimar la imagen original

VI. *Deconvolución*

Imágenes ruidosas y borrosas

VI. Deconvolución

Imágenes deconvolucionadas utilizando modelo a priori log

Software disponible en http://www.dbabacan.info/BDGSP.php

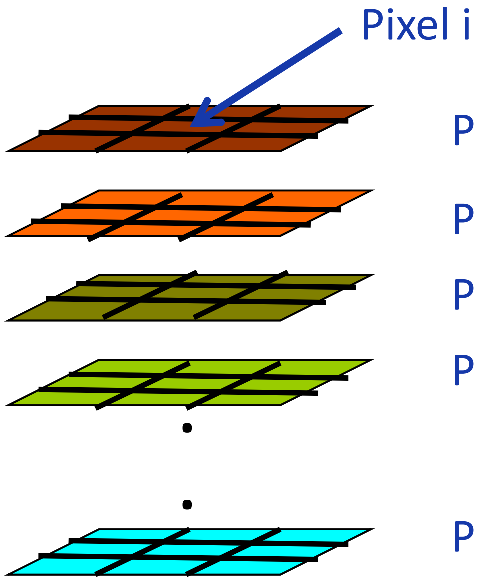
S.D. Babacan, R. Molina, M.N. Do, y A.K. Katsaggelos, "Blind deconvolution with general sparse image priors" in *European Conference on Computer Vision (ECCV)*, 341-355, 2012.

VI. Filtrado y Clasificación

- En problemas de clasificación supervisada, tenemos acceso a un conjunto de entrenamiento (N=nº. de muestras), y la etiqueta de cada muestra

$$\mathbf{Z} \in \mathbb{R}^{B \times N} \qquad\qquad \mathbf{y} \in \{0, 1\}^{1 \times N}$$

B es el número de bandas de la señal.



Pixel i y sus vecinos en la banda 1 $\mathbf{z}_{1,i}$

Pixel i y sus vecinos en la banda 2 $\mathbf{z}_{2,i}$

Pixel i y sus vecinos en la banda 3 $\mathbf{z}_{3,i}$

Pixel i y sus vecinos en la banda 4 $\mathbf{z}_{4,i}$

Pixel i y sus vecinos en la banda B $\mathbf{z}_{B,i}$

VI. Filtrado y Clasificación

- Entrenemos un clasificador y conjuntamente estimemos un banco de filtros para mejorar el calsificador

$$\mathbf{X} = \mathbf{A}\mathbf{Z} = \begin{bmatrix} \mathbf{a}_1^T & & \ldots & \\ & \mathbf{a}_2^T & \ldots & \\ \vdots & \vdots & \ddots & \vdots \\ & & \ldots & \mathbf{a}_B^T \end{bmatrix} \begin{bmatrix} \mathbf{z}_{1,1} & \mathbf{z}_{1,2} & \ldots & \mathbf{z}_{1,N} \\ \mathbf{z}_{2,1} & \mathbf{z}_{2,2} & \ldots & \mathbf{z}_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{z}_{B,1} & \mathbf{z}_{B,2} & \ldots & \mathbf{z}_{B,N} \end{bmatrix},$$

- Queremos construir el clasificador usando los rasgos filtrados en $\mathbf{X}$. También necesitamos calcular los coeficientes de los filtros en $\mathbf{A}$

VI. Filtrado y Clasificación

- El clasificador se construye usando Procesos Gaussianos

$$\mathrm{p}(\mathbf{y}|\mathbf{f}) = \prod_{i=1}^{N} [\boldsymbol{\sigma}(f_i)]^{y_i} [1 - \boldsymbol{\sigma}(f_i)]^{1-y_i}$$

donde $\boldsymbol{\sigma}$ representa la función sigmoidal y

$$\mathbf{f} \sim \mathcal{N}(\mu\mathbf{1}, \mathbf{C} = \gamma\mathbf{K}_{\mathbf{X}} + \sigma\mathbf{I})$$

$\mathbf{K}_{\mathbf{X}}$ es la matriz del núcleo

μ, γ, σ son parámetros a estimar

VI. Filtrado y Clasificación

- Para imponer $\mathbf{X}=\mathbf{A}\mathbf{Z}$, introducimos el siguiente modelo de pseudo observación

$$\mathrm{p}(\mathbf{X}|\mathbf{A},\mathbf{Z}) \propto \exp\left(-\frac{\beta}{2}\|\mathbf{X}-\mathbf{A}\mathbf{Z}\|_F^2\right)$$

cuando $\beta \rightarrow \infty$ tenemos la restricción $\mathbf{X}=\mathbf{A}\mathbf{Z}$

- Sobre los filtros usamos las siguientes distribuciones a priori:

$$\mathrm{p}(\mathbf{A}|\alpha) = \prod_{i=1}^{B} \mathrm{p}(\mathbf{a}_i|\alpha_i) = \prod_{i=1}^{B} \mathcal{N}(\mathbf{a}_i|\mathbf{0}, \alpha_i^{-1}\mathbf{I}_{k^2})$$

$$\mathrm{p}(\alpha) = \prod_{i=1}^{B} \mathrm{p}(\alpha_i) = \prod_{i=1}^{B} \Gamma(\alpha_i|a_i, b_i)$$

VI. Filtrado y Clasificación

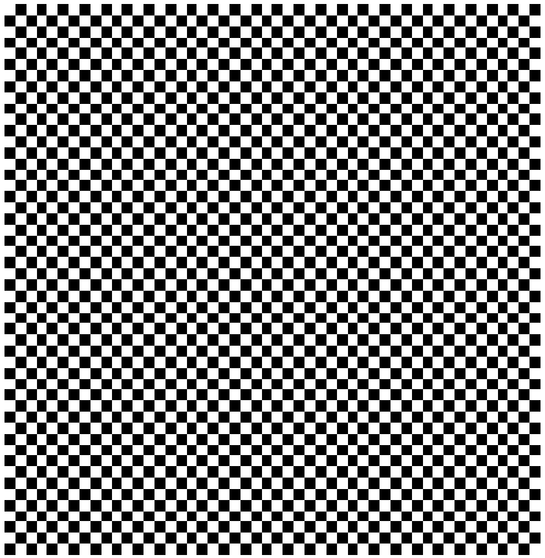
- Obtenemos la distribución conjunta:

$$\begin{aligned}\mathrm{p}(\mathbf{y},\mathbf{f},\mathbf{X},\mathbf{A},\alpha,\mu,\gamma,\sigma) = {} & \mathrm{p}(\mathbf{y}|\mathbf{f})\mathrm{p}(\mathbf{f}|\mathbf{X},\mu,\gamma,\sigma) \\ & \times \mathrm{p}(\mathbf{X}|\mathbf{A},\mathbf{Z})\mathrm{p}(\mathbf{A}|\alpha)\mathrm{p}(\alpha)\end{aligned}$$

- Usamos VB para realizar inferencia

VI. Filtrado y Clasificación

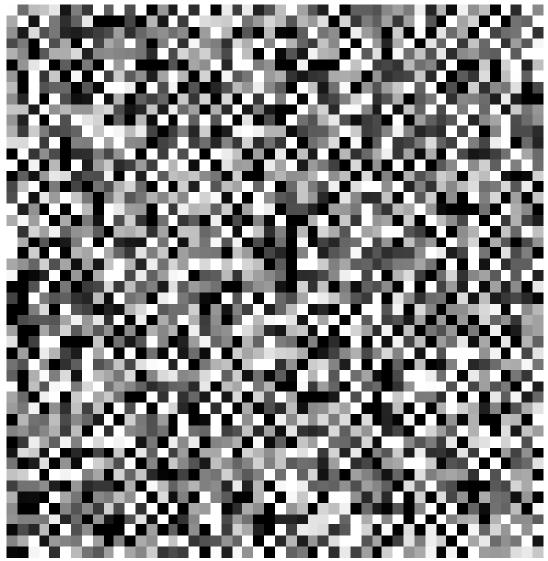
- Experimento sintético:
  Dado un tablero de ajedrez, generamos los observaciones usando



Tablero de ajedrez original

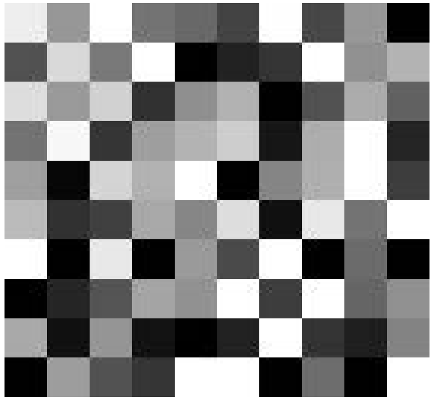
Cuadrados negros: realizaciones de una distribución gaussiana con media 0.25 y varianza 0.16

Cuadrados blancos: realizaciones de una distribución gaussiana con media 0.75 y varianza 0.16

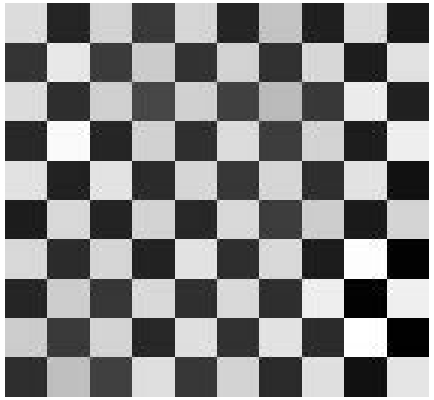


Observaciones

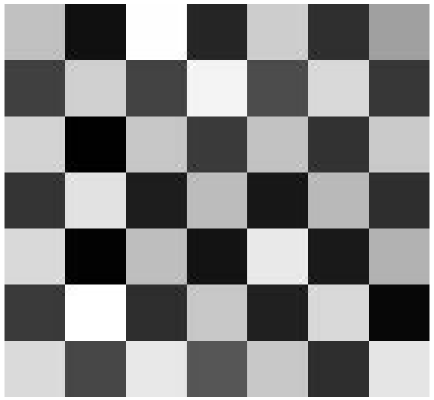
VI. Filtrado y Clasificación



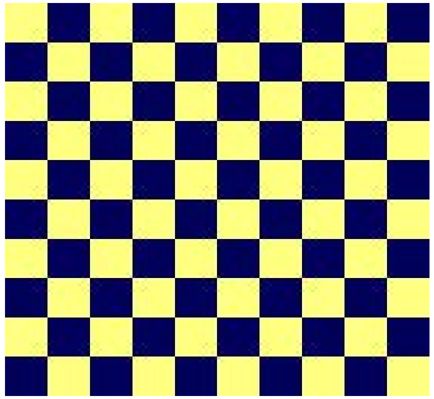
Observaciones



Observaciones filtradas



Coeficientes del filtro estimado



Clasificación

VI. Filtrado y Clasificación

- Base de datos real

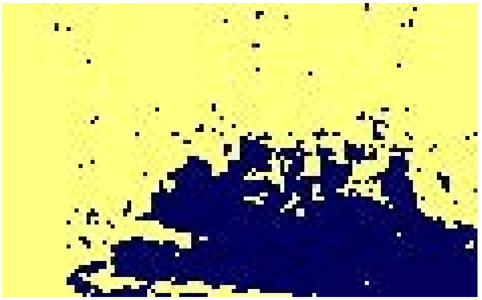
Imagen original

Imagen Landsat de Nápoles. Año 1995

El objetivo es clasificar los píxeles en las clases

Urbano contra No-Urbano

Clasificación verdadera



Clasificación sin filtrado

Clasificación con filtrado

P. Ruiz, J. Mateos, R. Molina, and A.K. Katsaggelos, "Learning Filters in Gaussian Process Classification Problems" in *IEEE International Conference on Image Processing*, Paris (France), October 2014.

Algunas referencias recientes (2014)

- L. Song, F. Jiang, Z. Shi, R. Molina, and A.K. Katsaggelos, “Dynamic Scene Understanding by Hierarchical Motion Pattern Mining”, IEEE Transactions on Intelligent Transportation Systems, vol. 15, 1273-1285, 2014.
- Z. Chen, S. D. Babacan, R. Molina, and A. K. Katsaggelos, “Variational Bayesian Methods For Multimedia Problems”, IEEE Transaction on Multimedia, vol. 16, 1000-10017, 2014.
- P. Ruiz, J. Mateos, G. Camps-Valls, R. Molina, and A.K. Katsaggelos, “Bayesian Active Remote Sensing Image Classification”, IEEE Transactions on Geoscience and Remote Sensing, vol. 52, 2186-2196, April 2014.
- Z. Chen, R. Molina, and A.K. Katsaggelos, “Automated Recovery of Compressed Sparse Signals From Smooth Background”, Signal Processing Letters, vol. 21, 1012-106, August 2014.
- S. Villena, M. Vega, R. Molina, and A.K. Katsaggelos, “A Non-stationary Image Prior Combination in Super-Resolution”, Digital Signal Processing, vol. 32, 1-10, September 2014.
- P. Ruiz, H. Madero-Orozco, J. Mateos, O.O. Vergara-Villegas, R. Molina, and A.K. Katsaggelos, “Combining Poisson Singular Integral and Total Variation prior models in Image Restoration”, Signal Processing, vol. 103, 296-308, October 2014.
- M. Luessi, S.D. Babacan, R. Molina, J.R. Booth, and A.K. Katsaggelos, “Variational Bayesian Causal Connectivity Analysis for fMRI”, Frontiers in Neuroinformatics, vol. doi: 10.3389/fninf.2014.00045, 2014.

Algunas referencias recientes

- Z. Chen, R. Molina, and A.K. Katsaggelos, "Recovery of correlated sparse signals from under-samples measurements" in *22th European Signal Processing Conference (EUSIPCO 2014)*, Lisbon (Portugal), 2014.
- M. Vega, R. Molina, and A.K. Katsaggelos, "Parameter Estimation in Bayesian Blind Deconvolution with Super Gaussian Image Priors" in *22th European Signal Processing Conference (EUSIPCO 2014)*, Lisbon (Portugal), September 2014.
- P. Ruiz, N. Pérez de la Blanca, R. Molina, and A. K. Katsaggelos, "Bayesian Classification and Active Learning Using Lp-Priors. Application to Image Segmentation" in *22th European Signal Processing Conference (EUSIPCO 2014)*, Lisbon (Portugal), September 2014.
- P. Ruiz, J. Mateos, R. Molina, and A.K. Katsaggelos, "Learning Filters in Gaussian Process Classification Problems" in*IEEE International Conference on Image Processing*, Paris (France), October 2014.
- X. Zhou, R. Molina, F. Zhou, and A.K. Katsaggelos, "Fast iteratively reweighted least squares for Lp regularized image deconvolution and reconstruction" in *IEEE International Conference on Image Processing*, Paris (France), October 2014.

Visita http://decsai.ugr.es/vip/

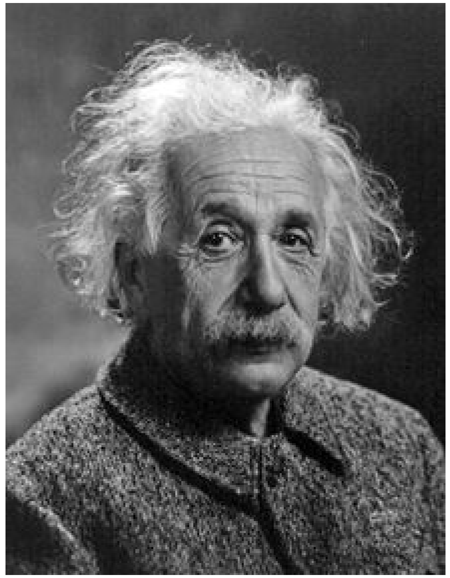
Resumen

I ♥ KL(Q‖P)

Imagen de la Shannon lecture de Sergio Verdú en 2007

Ya que siempre debemos intentar

Make things as simple as possible, but not simpler.



Gracias