Análisis del artículo

Mean Shift: A Robust Approach toward Feature Space Analysis

Dorin Comaniciu - Peter Meer

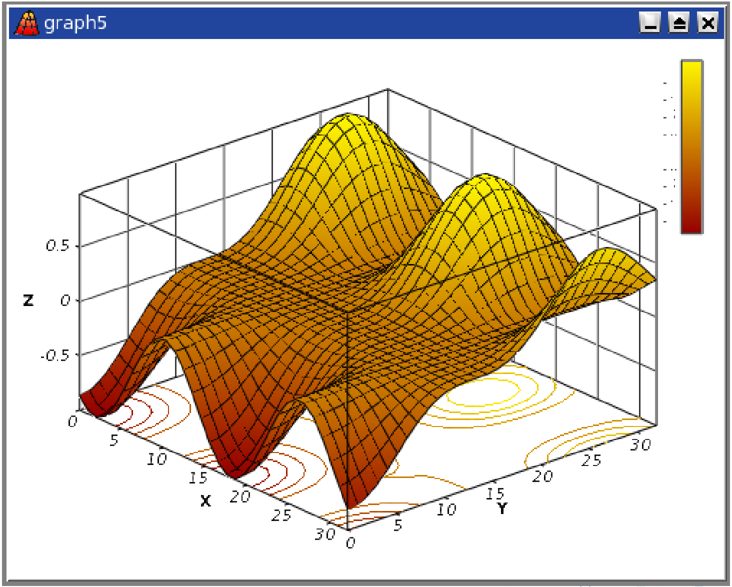
—

R. Omar Chávez Garcia - Pável Herrera Domínguez

25 de mayo de 2009

Conceptos básicos

Mínimos y máximos



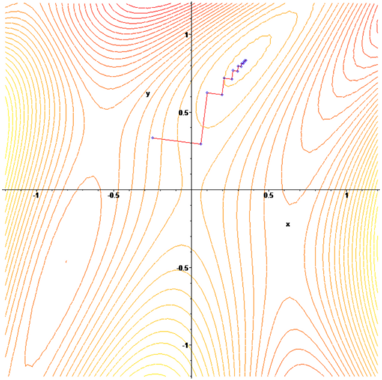
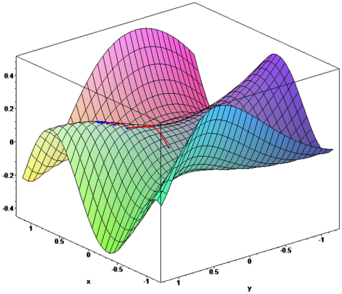
Gradientes

- Máximo o mínimo $\nabla f(x) = 0$

Gradientes

- Máximo o mínimo $\nabla f(x) = 0$
- Descenso/Ascenso de gradiente: $x_{t+1} = x_t + \lambda \nabla f(x_t)$

Gradientes

- Máximo o mínimo $\nabla f(x) = 0$
- Descenso/Ascenso de gradiente: $x_{t+1} = x_t + \lambda \nabla f(x_t)$
- 


Espacio de características

Es un espacio abstracto (n-dimensional) en el cual una muestra es representada como un punto. La dimensión del espacio está determinada por el número de características utilizadas para representar la muestra.

- Permite una representación fiable de los datos de entrada.
- El proceso de extracción de características es controlado por pocos parámetros.
- Su naturaleza es dependiente de la aplicación.

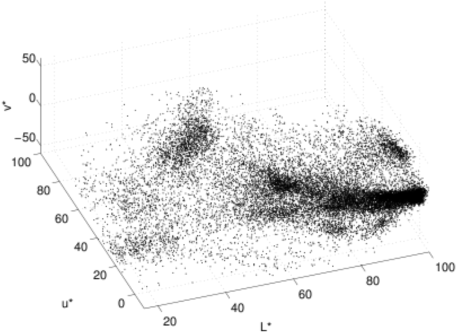
Análisis del espacio de características

El objetivo del análisis del espacio de características es el delinear los grupos representados como regiones densas de puntos(características importantes) en el espacio de características.

- La evidencia de la presencia de una característica importante es colocada junta.
- Características con poca concentración en el espacio de características a pesar de ser importantes para la aplicación, no serán detectadas.
- El análisis del espacio de características es independiente de la aplicación.
- Métodos que necesitan información a priori sobre el número de grupos o la forma de estos, no son capaces de manejar la complejidad de un espacio de características real.

Ejemplo

(a)



(b)

¿Qué es un Kernel? (1)

Representación común: *Suponga que tenemos un conjunto de objetos* $S = x_1, x_2, ..., x_n$ *y una representación* $\phi(S) = \phi(x_1), \phi(x_2), ..., \phi(x_n)$.
Representación basada en kernels: *Suponga que en vez de tener una representación* $\phi : X \rightarrow F$, *tenemos una representación* $k : X \times X \rightarrow R$, *con lo que el conjunto de entrada* S *queda representado por una matriz real* K *de dimensiones* $n \times n$, *donde* $K_{i,j} = k(x_i, x_j)$.

- La representación con kernels ofrece modularidad en el análisis de los datos ya que separa la etapa de representación del algoritmo de análisis.
- Se pueden utilizar algoritmos genéricos para el manejo de matrices de números reales independientemente de la función k que se utilice y el tipo de datos del dominio del problema.
- Algoritmos sencillos de análisis pueden encontrar relaciones complejas entre objetos de forma eficiente.

¿Qué es un Kernel? (2)

Definición (función kernel)

Una función kernel es una función $k : X \times X \rightarrow R$, que asigna a cada par de objetos del espacio de entrada, X, un valor real correspondiente al producto escalar de las imágenes de dichos objetos en un espacio, F, que se denomina espacio de características, es decir:
$k(x, y) =< \phi(x), \phi(y) >$;

- El objetivo de una función kernel es el de expresar en alguna forma la similitud entre dos objetos.

Tipos de Kernel

- kernel lineal: $k(x,y) =< x, y >$
- kernel polinómico: $k(x,y) = (< x, y >)^p$
- kernel gaussiano: $k(x,y) = e^{||x-y||^2/2\sigma^2}$
- kernel multivariado: $k(x) = |H|^{1/2} K(H^{1/2} x)$

Métodos paramétricos y no-paramétricos

- Existen espacios de características que solo pueden ser analizados por métodos no-paramétricos dado que estos no consideran supocisiones sobre el espacio.
- Los métodos no-paraméticos se pueden clasificar en:
  - Agrupamiento jerárquico. Unen o dividen conjuntos de datos según una medida de proximidad. Son computacionalmente costosos y no es claro definir un criterio de paro.
  - Estimación de densidad. Considerar al espacio de características como una función de densidad de probabilidad empírica (p.d.f.) del parámetro representado.
    Regiones densas en el espacio de características corresponden a máximos locales de la p.d.f., es decir, las *modas* de la densidad desconocida.
    Cuando la posición de la moda es encontrada el grupo asociado a ella es delineado según la estructura local del espacio de características.

Mean Shift

Procedimiento Mean Shift (1)

Definición (Mean Shift)

Sea $S \subset X$ *un conjunto finito (los datos muestra) en el espacio n-dimensional. Sea* K *un kernel y* $w : S \rightarrow (0, \infty)$ *una función de pesos. La muestra promedio con kernel* K *en* $x \epsilon X$ *es definida como:*

$$m(x) = \frac{\sum_{s \epsilon S} K(s-x)w(s)s}{\sum_{s \epsilon S} K(s-x)w(s)}$$

La diferencia $m(x) - x$ *es conocida como mean shift.*

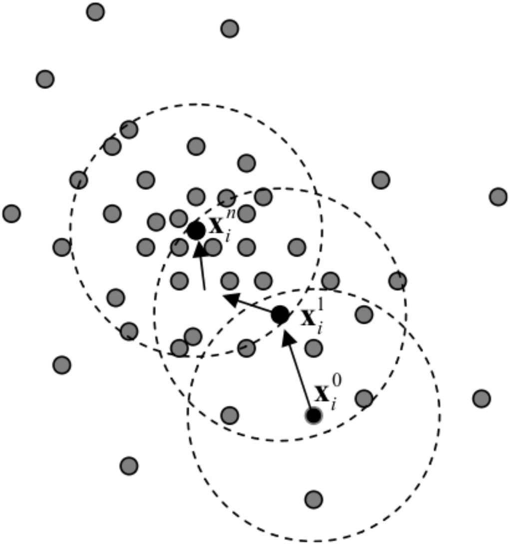
Procedimiento Mean Shift (2)

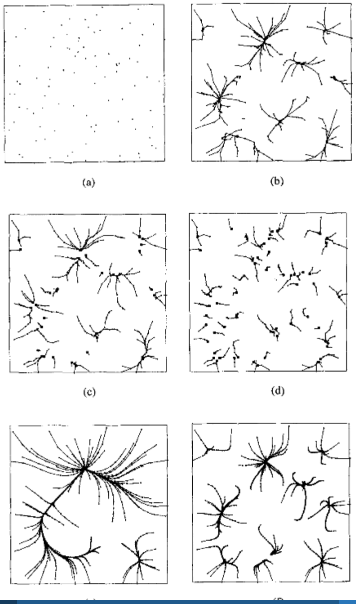
Definición (Algoritmo Mean Shift)

Sea $T \subset X$ un conjunto finito (centros de los grupos) en el espacio n-dimensional. La evolución de T en la forma de iteraciones $T \leftarrow m(T)$ con $m(T) = m(t); t \in T$, es conocida como el algoritmo mean shift. El algoritmo se detiene cuando alcanza un punto fijo($m(T)=T$).

Para cada $t \epsilon T$ hay una secuencia $t, m(t), m(m(t)), ...,$ que es llamada trayectoria de t.

Procedimiento Mean Shift(3)




(a)
(b)
(c)
(d)

Estimación de la densidad del gradiente

- Densidad se estima como (ventana de Parzen):

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n} K_H(x - x_i)$$

$$\hat{f}(x) = \frac{1}{nh^d}\sum_{i=1}^{n} K(x - x_i), h > 0$$

Estimación de la densidad del gradiente

- Densidad se estima como (ventana de Parzen):

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n} K_H(x - x_i)$$

$$\hat{f}(x) = \frac{1}{nh^d}\sum_{i=1}^{n} K(x - x_i), h > 0$$

- Cambiando de notación:

$$\hat{f}_{h,K}(x) = \frac{c_{k,d}}{nh^d}\sum_{i=1}^{n} k\left(\left\|\frac{x - x_i}{h}\right\|^2\right), h > 0$$

Estimación de la densidad del gradiente (2)

- Gradiente:

$$\nabla \hat{f}_{h,K}(x) = \frac{2 * c_{k,d}}{nh^{d+2}} \sum_{i=1}^{n} (x - x_i) k' \left(\left\| \frac{x - x_i}{h} \right\|^2 \right)$$

Estimación de la densidad del gradiente (2)

- Gradiente:

$$\nabla \hat{f}_{h,K}(x) = \frac{2 * c_{k,d}}{nh^{d+2}} \sum_{i=1}^{n} (x - x_i) k' \left(\left\| \frac{x - x_i}{h} \right\|^2 \right)$$

- Cambio de notación

$$\nabla \hat{f}_{h,K}(x) = \frac{2 * c_{k,d}}{nh^{d+2}} \sum_{i=1}^{n} (x - x_i) g \left(\left\| \frac{x - x_i}{h} \right\|^2 \right) =$$

Estimación de la densidad del gradiente (2)

- Gradiente:

$$\nabla \hat{f}_{h,K}(x) = \frac{2 * c_{k,d}}{nh^{d+2}} \sum_{i=1}^{n} (x - x_i) k' \left(\left\| \frac{x - x_i}{h} \right\|^2 \right)$$

- Cambio de notación

$$\nabla \hat{f}_{h,K}(x) = \frac{2 * c_{k,d}}{nh^{d+2}} \sum_{i=1}^{n} (x - x_i) g \left(\left\| \frac{x - x_i}{h} \right\|^2 \right) =$$

- Acomodando terminos

$$\frac{2 * c_{k,d}}{nh^{d+2}} \sum_{i=1}^{n} g \left(\left\| \frac{x - x_i}{h} \right\|^2 \right) \left[\frac{\sum_{i=1}^{n} x_i g \left(\left\| \frac{x - x_i}{h} \right\|^2 \right)}{\sum_{i=1}^{n} g \left(\left\| \frac{x - x_i}{h} \right\|^2 \right)} - x \right]$$

Estimación de la densidad del gradiente (2)

- Mean shift:

$$m_{h,G}(x) = \left[\frac{\sum_{i=1}^{n} x_i g\left(\left\| \frac{x - x_i}{h} \right\|^2 \right)}{\sum_{i=1}^{n} g\left(\left\| \frac{x - x_i}{h} \right\|^2 \right)} - x \right]$$

Estimación de la densidad del gradiente (2)

- Mean shift:

$$m_{h,G}(x) = \left[\frac{\sum_{i=1}^{n} x_i g\left(\left\| \frac{x - x_i}{h} \right\|^2 \right)}{\sum_{i=1}^{n} g\left(\left\| \frac{x - x_i}{h} \right\|^2 \right)} - x \right]$$

- Estimación de densidad:

$$\hat{f}_{h,G} = \frac{c_{k,d}}{nh^{d+2}} \sum_{i=1}^{n} g\left(\left\| \frac{x - x_i}{h} \right\|^2 \right)$$

Estimación de la densidad del gradiente (2)

- Mean shift:

$$m_{h,G}(x) = \left[\frac{\sum_{i=1}^{n} x_i g\left(\left\| \frac{x - x_i}{h} \right\|^2 \right)}{\sum_{i=1}^{n} g\left(\left\| \frac{x - x_i}{h} \right\|^2 \right)} - x \right]$$

- Estimación de densidad:

$$\hat{f}_{h,G} = \frac{c_{k,d}}{nh^{d+2}} \sum_{i=1}^{n} g\left(\left\| \frac{x - x_i}{h} \right\|^2 \right)$$

- Mean shift interpretación.

$$m_{h,G}(x) = \frac{1}{2} h^2 c_{k,d} \frac{\nabla \hat{f}_{h,K}(x)}{\hat{f}_{h,G}(x)}$$

- Iteración del mean shift

$$y_{t+1} = y_t + \left[\frac{\sum_{i=1}^{n} x_i g\left(\left\| \frac{y_t - x_i}{h} \right\|^2 \right)}{\sum_{i=1}^{n} g\left(\left\| \frac{y_t - x_i}{h} \right\|^2 \right)} - y_t \right]$$

Identificación de modas

Las trayectorias de este método de descenso de gradiente son atraidas por un máximo local si son puntos estacionarios únicos(para una vecindad pequeña).

- Hacer el proceso de mean shift para encontrar los puntos estacionarios.
- Podar estos puntos y solo retener los máximos locales.

Propiedad de trayectoria suave

Si se usa un kernel gaussiano, la trayectoria que sigue el método de mean shift es una trayectoria suave, esto es el angulo entre dos vectores mean shift siempre es menor que 90 grados.

Aplicaciones

Mean Shift para el análisis del espacio de características

- El procedimiento Mean Shift es una herramienta muy versátil para el análisis del espacio de características y provee soluciones fiables para tareas de visión computacional.
- La calidad para la localización de modas en la "colina" en la que fue inicializado hace a este método ideal para analizar espacios reales en los que el número de cluster y sus formas no son "predecibles".
- Es necesario establecer dos puntos importantes:
  - Una métrica para el espacio de características.
  - La forma del kernel a utilizar.
- Encontrar la forma de un grupo particular no representa un problema para el procedimiento Mean Shift ya que es resultado de la búsqueda de la moda del grupo. La forma la determinan todos los puntos visitados por los procedimientos Mean Shift realizados que llevaron a la convergencia de la moda.

Selección del *bandwidth*

Existen 4 diferentes técnicas para la selección del parámetro *bandwidth* h utilizado en la función kernel paral el procedimiento Mean Shift:

- Selección estadística. Es el *bandwidth* que alcanza el mejor compromiso entre el sesgo y la varianza del estimador sobre todos los $x \epsilon R^d$.
- Por estabilidad de la descomposición. El *bandwidth* es el centro del rango mas grande sobre el cual se obtuvo el mismo número de grupos para los datos dados.
- *Bandwidth* máximo. Es el mejor *bandwidth* que maximiza una función objetivo que expresa la calidad de descomposición.
- Información de alto nivel. Como en muchos de los casos la descomposición depende de la tarea a realizar, se puede tomar el *bandwidth* de la información dada por el usuario o por niveles altos de la tarea.

Filtros

1. Inicializa j=1 y $y_{i,1} = x_i$

Filtros

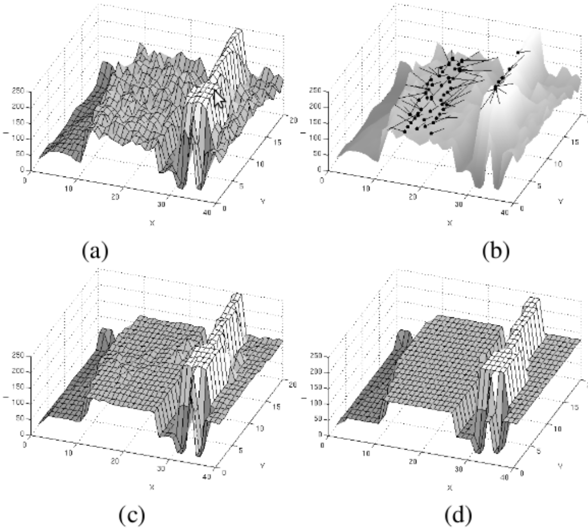
1. Inicializa j=1 y $y_{i,1} = x_i$
2. Calcula $y_{i,j+1}$ usando la iteración del mean shift hasta converger a $y = y_{i,c}$

Filtros

1. Inicializa j=1 y $y_{i,1} = x_i$
2. Calcula $y_{i,j+1}$ usando la iteración del mean shift hasta converger a $y = y_{i,c}$
3. Asignar $z_i = (x_i^s, y_{i,c}^r)$

Filtros

Graficas



Segmentación

1. Correr el proceso para filtrar usando mean shift y guardar toda la información sobre la convergencia del método.

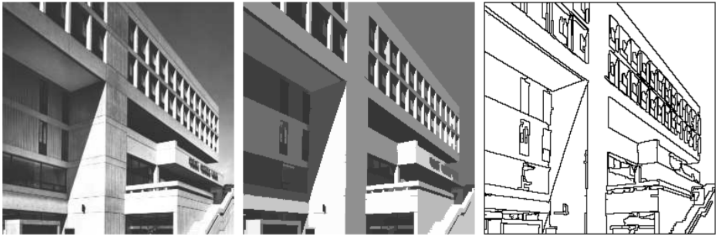
Segmentación

1. Correr el proceso para filtrar usando mean shift y guardar toda la información sobre la convergencia del método.
2. Construye los clusters $\{C_p\}_{p=1\ldots m}$ agrupando todos los z_i que estan a una distancia h_s en el dominio espacial, y h_r en el dominio del rango.

Segmentación

1. Correr el proceso para filtrar usando mean shift y guardar toda la información sobre la convergencia del método.
2. Construye los clusters $\{C_p\}_{p=1...m}$ agrupando todos los z_i que estan a una distancia h_s en el dominio espacial, y h_r en el dominio del rango.
3. Para cada etiqueta $i = 1, ..., n$ asigna $L_i = \{p | z_i \in C_p\}$

Segmentación



(a) (b) (c)

¿Preguntas o comentarios?

Recursos

- Librerás y aplicaciones
  - Librería en C++: http://www.caip.rutgers.edu/riul/research/code/EDISON/index.html
  - Librería en Matlab: http://www.lems.brown.edu/vision/project/robby/
  - Aplicación para *tracking*: http://www.cs.bilkent.edu.tr/~ismaila/MUSCLE/MSTracker.htm
- Cursos y tutoriales
  - Curso de visión: http://www.wisdom.weizmann.ac.il/~deniss/vision_spring04/lectures_full.html
  - Ponencia sobre k-means, mean shift,SVM : http://videolectures.net/mlss06tw_meila_co/
  - Diapositivas sobre mean shift: http://www.wisdom.weizmann.ac.il/~deniss/vision_spring04/files/mean_shift/mean_shift.ppt

Referencias

- *Mean Shift: A Robust Approach toward Feature Space Analysis*. Dorin Comaniciu, Peter Meer. Siemens Corporate Research - Electrical and Computer Engineering Department Rutgers University.
- *Mean Shift, Mode Seeking, and Clustering*. Yizong Cheng.
- *Mean Shift Analysis and Applications*. Dorin Comaniciu, Peter Meer. Siemens Corporate Research - Electrical and Computer Engineering Department Rutgers University.
- *Aprendizaje Automático: conceptos básicos y avanzados*. Basilio Sierra Araujo, coordinador. Pearson - Prentice Hall, 2006.