Posibles *features* de nodos y como clasifican no supervisadamente

# Introducción

En este informe pretendo exponeros a todos algunas de las *features* obtenidas a partir de los nodos que pueden ayudar en la clasificación sobre malignidad. Son *features* intuitivas y/o utilizadas en bibliografía que deberían estar relacionadas con aspectos más subjetivos, como la espicularidad o la calcificación, decisivos a la hora de valorar la malignidad de un nódulo.

No se discute aquí nada sobre **si un nódulo es realmente un nódulo o no.** Todos los nódulos que se presentan tienen un score superior a 0.8, bien se hayan obtenido con el ground truth de LUNA o bien mediante el score que da la red de DL de DSB (que no sé cuál es).

Lo que se ha hecho es definir una serie de *features* o grupos de *features*  que se analizan conjuntamente y se clasifican en dos grupos mediante kmeans y se visualiza con PCA en un espacio tridimensional, excepto si es una sola *feature* que se muestra mediante dos histogramas, uno para cada clase de separación. Luego, se ha muestreado para cada feature o *features* o grupo de *features*  tres nódulos de cada clase. En las imágenes es sobre todo importante fijarse en la primera columna (escala de grises) y en la última (máscara del nódulo, es decir, forma). Por otro lado, se ha cogido una muestra aleatoria de cada uno, para ver si clasifica o no, pero junto con este documento incluyo un csv con todas las features por si alguien quiere ver la pauta general de cada una de ellas en la muestra.

Salvo en el caso de la solidez, que son imágenes repetidas puestas a posteriori sin muestrear realmente, las imágenes de las tablas deben leerse, por ejemplo en la primera tabla, así.

|  |  |
| --- | --- |
| #1 | #2 |
| #3 | #4 |
| #5 | #6 |

# Listado de nódulos

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **patientid** | **nslice** | **x** | **y** | **diameter** | **#** |
| luna\_141511313712034597336182402384.npz | 100 | 331 | 317 | 4,22E+06 | 1 |
| dsb\_48fac79fead32f2b10e37752cb077af6.npz | 71 | 334 | 340 | 8,44E+06 | 2 |
| dsb\_43933b4021d93dd64854f318656c7d1e.npz | 133 | 310 | 213 | 4,37E+06 | 3 |
| dsb\_95edad4544e59e37188dd33db40cecb7.npz | 73 | 230 | 394 | 4,07E+06 | 4 |
| dsb\_b4581f4f4cad292b5a013d35d1c39f24.npz | 110 | 328 | 145 | 9,84E+06 | 5 |
| dsb\_1cf8e778167d20bf769669b4be96592b.npz | 96 | 262 | 100 | 3,74E+06 | 6 |
| dsb\_0482c444ac838adc5aa00d1064c976c1.npz | 83 | 396 | 177 | 4,79E+06 | 7 |
| dsb\_95edad4544e59e37188dd33db40cecb7.npz | 84 | 293 | 205 | 3,91E+06 | 8 |
| luna\_242624386080831911167122628616.npz | 57 | 318 | 96 | 5,17E+06 | 9 |
| dsb\_48fac79fead32f2b10e37752cb077af6.npz | 85 | 343 | 125 | 5,05E+06 | 10 |
| dsb\_e42815372aa308f5943847ad06f529de.npz | 140 | 417 | 342 | 5,64E+06 | 11 |
| dsb\_f599c9bb0bd19126e13fba8b63d2de41.npz | 52 | 339 | 170 | 4,79E+06 | 12 |
| dsb\_f228862e9f9c6ad7993ded4e088c5050.npz | 110 | 312 | 217 | 1,41E+07 | 13 |
| luna\_159665703190517688573100822213.npz | 70 | 315 | 165 | 3,57E+06 | 14 |
| dsb\_c6b722b00948587dd89a548f9639f447.npz | 90 | 231 | 199 | 5,64E+06 | 15 |
| dsb\_daee58dd09c005fe41ca370c2a2fdfa9.npz | 59 | 344 | 208 | 6,68E+06 | 16 |
| dsb\_9de48cf43611478ffc1fef051b75dc8c.npz | 67 | 293 | 181 | 9,58E+06 | 17 |
| luna\_171919524048654494439256263785.npz | 90 | 239 | 223 | 4,37E+06 | 18 |
| luna\_803808126682275425758092691689.npz | 42 | 311 | 319 | 8,06E+06 | 19 |
| dsb\_f228862e9f9c6ad7993ded4e088c5050.npz | 120 | 345 | 327 | 6,58E+06 | 20 |
| luna\_803808126682275425758092691689.npz | 101 | 260 | 169 | 5,17E+06 | 21 |
| luna\_286647622786041008124419915089.npz | 61 | 257 | 105 | 9,51E+06 | 22 |
| luna\_211956804948320236390242845468.npz | 82 | 311 | 343 | 3,57E+06 | 23 |
| luna\_276351267409869539593937734609.npz | 128 | 324 | 351 | 3,57E+06 | 24 |
| dsb\_67ccc2ecd7716e1cab05caaedea99f4b.npz | 60 | 332 | 214 | 4,07E+06 | 25 |
| dsb\_8c6f2ce60479f73c257173911aff3453.npz | 103 | 309 | 432 | 1,04E+07 | 26 |
| dsb\_2619ed1e4eca954af4dcbc4436ef8467.npz | 62 | 296 | 108 | 5,53E+06 | 27 |
| dsb\_e4ff18b33b7110a64f497e177102f23d.npz | 63 | 242 | 159 | 7,90E+06 | 28 |
| dsb\_2619ed1e4eca954af4dcbc4436ef8467.npz | 65 | 367 | 182 | 1,37E+07 | 29 |
| dsb\_7f096cdfbc2fe03ec7f779278416a78c.npz | 84 | 310 | 354 | 7,31E+06 | 30 |
| dsb\_daee58dd09c005fe41ca370c2a2fdfa9.npz | 45 | 337 | 346 | 1,19E+07 | 31 |
| dsb\_6a08bc7a3156caf9b07bbb1f9b2861d2.npz | 67 | 257 | 158 | 1,08E+07 | 32 |
| luna\_242624386080831911167122628616.npz | 58 | 415 | 331 | 3,91E+06 | 33 |
| dsb\_4b6b53281fe1fe797e181a5287df0f9a.npz | 85 | 164 | 383 | 6,28E+06 | 34 |
| dsb\_de04fbf5e6c2389f0d039398bdcda971.npz | 104 | 334 | 384 | 4,22E+06 | 35 |
| dsb\_4b6b53281fe1fe797e181a5287df0f9a.npz | 72 | 346 | 389 | 5,75E+06 | 36 |

# Features

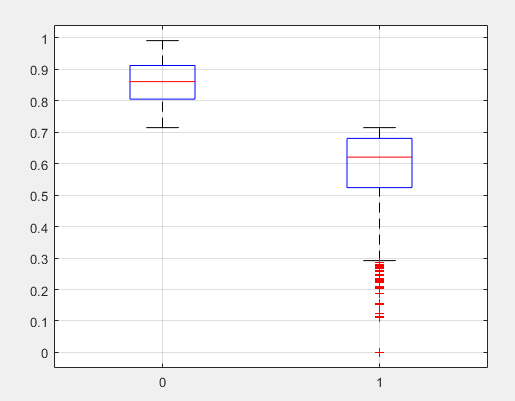
## Esbeltez

### Definición

Se define como

Siendo y el menor y mayor autovalor de la matriz de covarianzas de la imagen. A priori, cuanto más esbelto el nódulo, menos circular. No todos los nódulos esbeltos son espiculados pero los nódulos espiculados serán más esbeltos.

### Clases del kmeans



### Ejemplos de cada clase

|  |  |
| --- | --- |
|  |  |

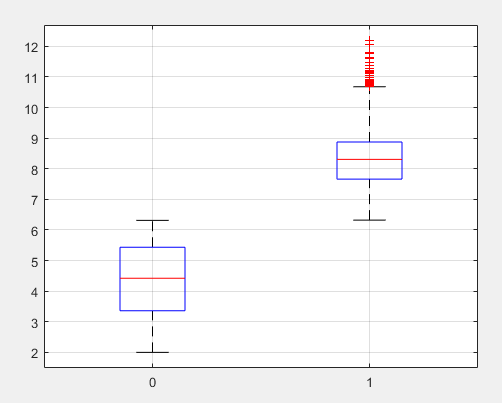
## Radio equivalente

### Definición

Se define como

Siendo el área el perímetro. Aquí hay dos efectos: el efecto del tamaño ( y el efecto de la ( irregularidad. La malignidad por lo tanto sería difícil de discriminar con este parámetro pero quizás en combinación con otros, ayude.

### Clases del kmeans



### Ejemplos de cada clase

|  |  |
| --- | --- |
|  |  |

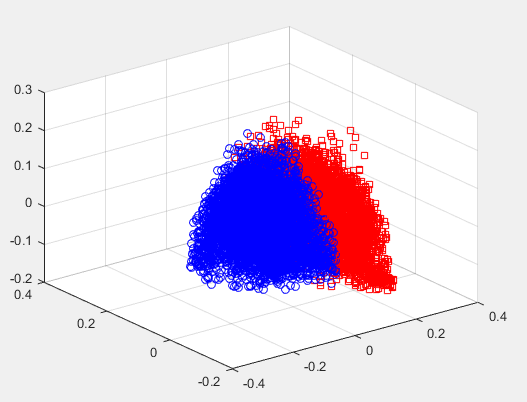
## Histogram of gradients

### Definición

<https://en.wikipedia.org/wiki/Histogram_of_oriented_gradients>

En teoría codifican la forma a un nivel menos básico que el mapa de píxeles. No tengo mucha intuición sobre como ayudarían a priori en la detección de malignidad (espicularidad o calcificación).

### Clases del kmeans



### Ejemplos de cada clase

|  |  |
| --- | --- |
|  |  |

## HU moments

### Definición

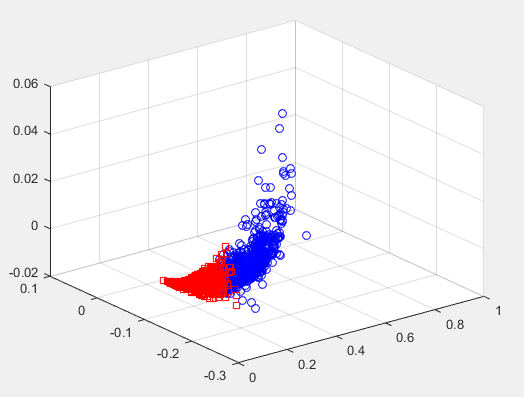
<http://www.sci.utah.edu/~gerig/CS7960-S2010/handouts/Hu.pdf>

Referencias de donde lo usan para clasificación de nódulos en cáncer de pulmón:

<http://www.sciencedirect.com/science/article/pii/S1361841513001722>

Tampoco tengo mucha intuición, pero son invariantes bajo rotaciones y semejanzas.

### Clases del kmeans



### Ejemplos de cada clase

|  |  |
| --- | --- |
|  |  |

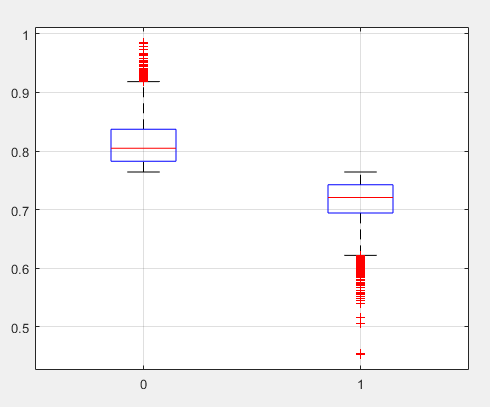
## Homogeneidad

### Definición

Se define como

Donde es la intensidad media del nódulo y es la máxima. A priori, cuanto mayor homogeneidad tenga un nódulo, más calcificado estará y por lo tanto menos probabilidades de que sea maligno.

### Clases del kmeans



### Ejemplos de cada clase

|  |  |
| --- | --- |
|  |  |

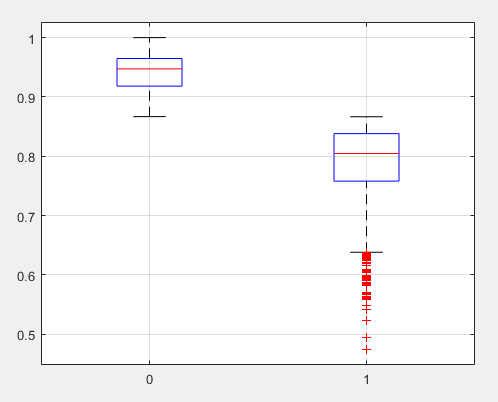
## Solidez

### Definición

Se define como

Siendo el área y el área de la envolvente conexa. En principio, cuanto más pequeño sea este valor, más irregular es la forma, que es una posible manifestación de la espicularidad.

### Clases del kmeans



### Ejemplos de cada clase[[1]](#footnote-1)

|  |  |
| --- | --- |
|  |  |

## Local binary patterns

### Definición

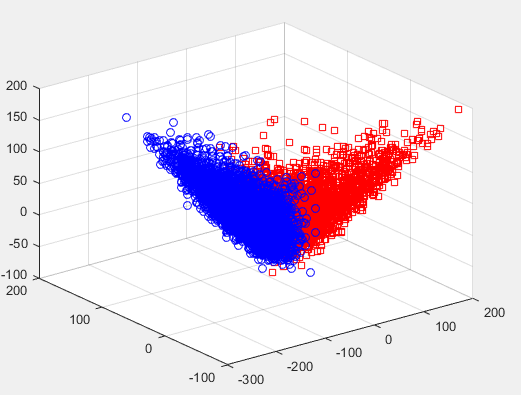
<https://en.wikipedia.org/wiki/Local_binary_patterns>

Referencias en las que se usa para identificación o clasificación de nódulos:

<http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=5872380>

<https://link.springer.com/chapter/10.1007/978-3-540-74260-9_93>

### Clases del kmeans



### Ejemplos de cada clase

|  |  |
| --- | --- |
|  |  |

1. Se recuerda que estas imágenes están repetidas, corresponden a las tres de mayor solidez y a las tres de menos solidez, consultando el archivo csc. [↑](#footnote-ref-1)