
Breast Cancer Wisconsin (Diagnostic)

Aplicando el teorema de Bayes

Carlos Catalán - 3 de marzo de 2020



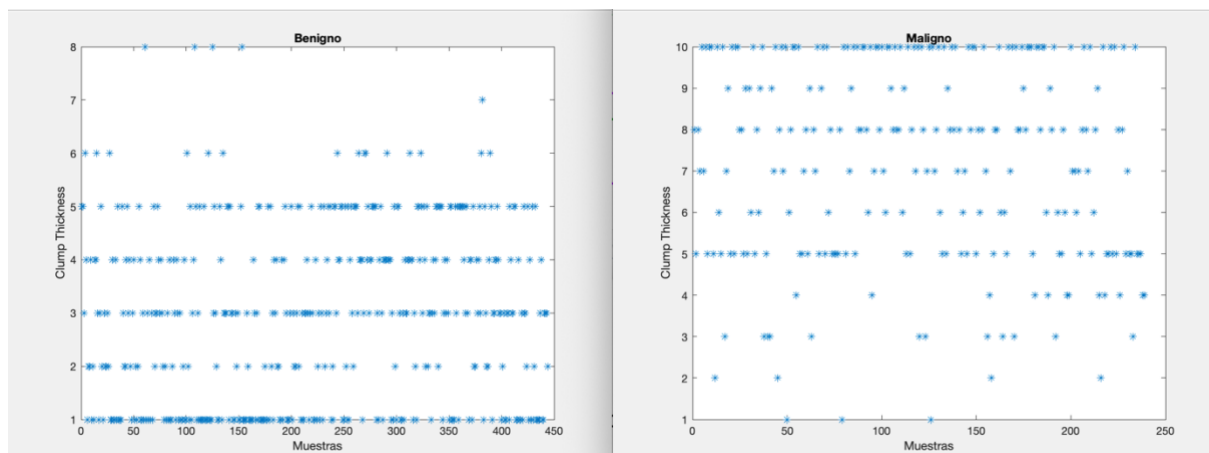
Base de datos: [https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+\(Diagnostic\)?fbclid=IwAR0U7thoOSK4ZKLNnkXSQmyRkyAuwCTlnV6z9vva2JSI5hX--y-kqdBBgg4](https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic)?fbclid=IwAR0U7thoOSK4ZKLNnkXSQmyRkyAuwCTlnV6z9vva2JSI5hX--y-kqdBBgg4)

Reporte

Observando el comportamiento de los datos podemos observar que hay una diferencia notable en las dos principales categorías (Benigno o Maligno). Entre la información mas destacable que se puede resaltar de ahí, se encuentra la siguiente:

Numero de muestras totales: 683
Benignos: 444
Malignos: 239

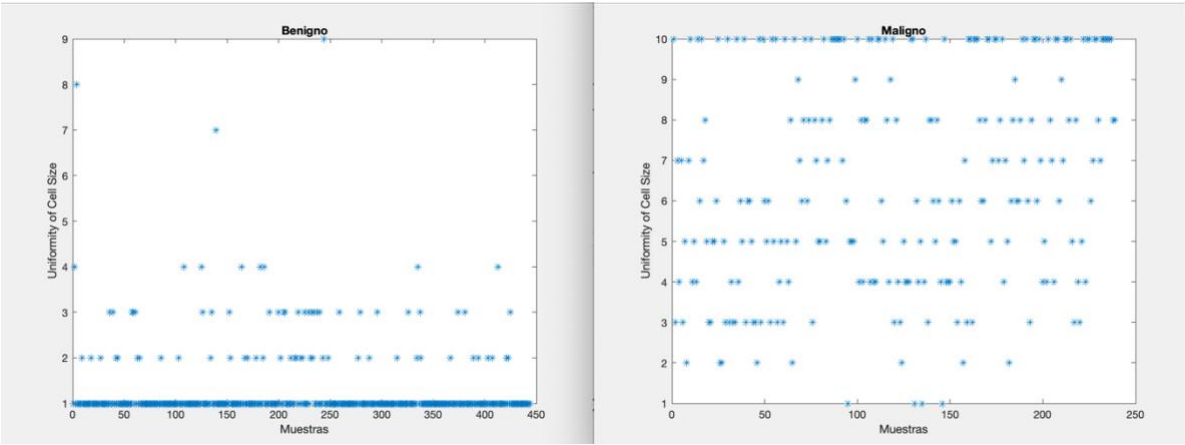
Clump Thickness: 1-10 (Grosor del grupo de células)



Media Benigno:	Desviación Estándar Benigno:	Media Maligno:	Desviación Estándar Maligno:
2.9640	1.6727	7.1883	2.4379

Se puede observar que en el grosor del grupo de células hay una diferencia notable entre los diferentes tipos de tumores, otra observación es que la gran concentración de tumores benignos se encuentran abajo del grosor 5, por lo que tener un tamaño mayor a este se puede interpretar como un factor de riesgo a tomar en cuenta.

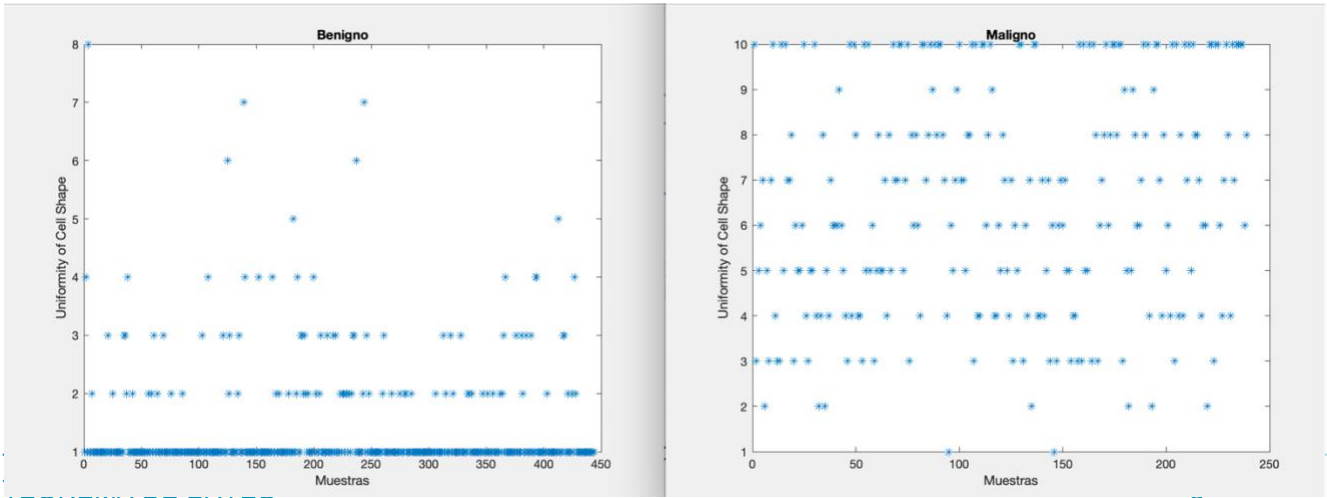
Uniformity of Cell Size: 1-10 (Uniformidad del tamaño de la célula)



Media Benigno:	Desviación Estándar Benigno:	Media Maligno:	Desviación Estándar Maligno:
1.3063	0.8557	6.5774	2.7242

Cómo se observa en esta característica, después de la uniformidad de tamaño 1, se encuentra casi toda la concertación de tumores malignos, habiendo una probabilidad muy grande de que el tumor sea maligno con un tamaño mayor a 1.

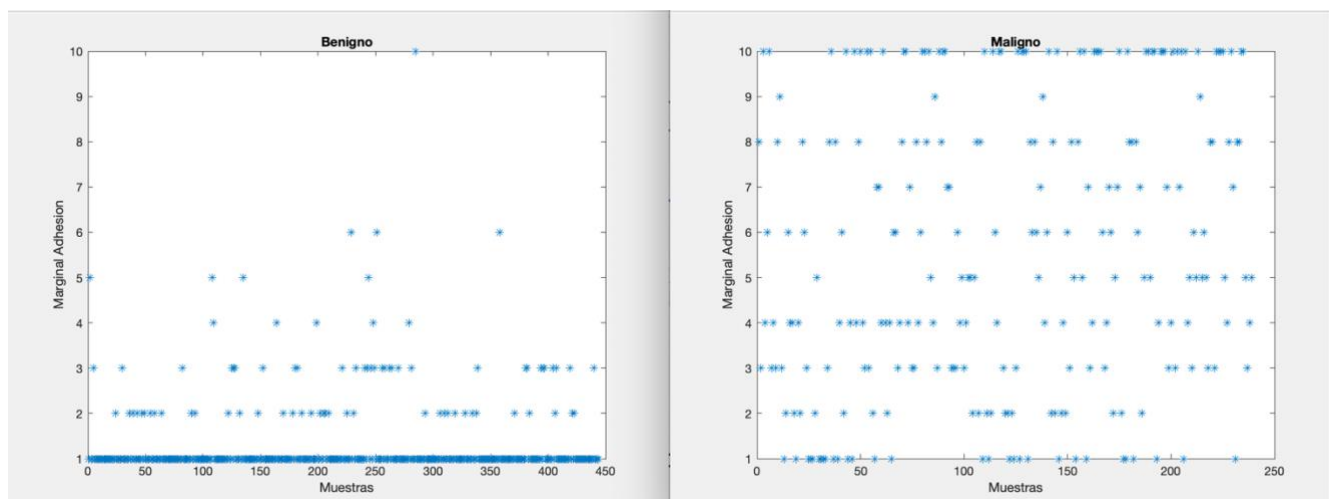
Uniformity of Cell Shape: 1-10 (Uniformidad de la forma de la célula)



Media Benigno:	Desviación Estándar Benigno:	Media Maligno:	Desviación Estándar Maligno:
1.4144	0.9570	6.5607	2.5691

Aquí se puede observar algo muy parecido con lo uniforme del tamaño de la célula, como se aprecia la forma de esta, también es un factor clave a la hora de diagnosticar un tumor maligno.

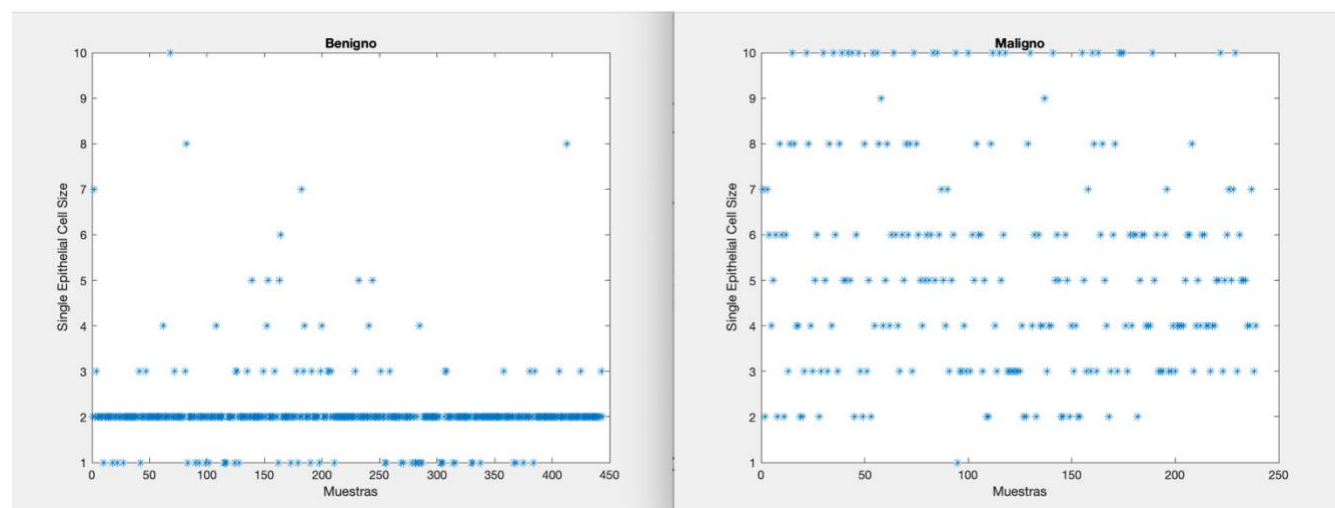
Marginal Adhesion: 1-10 (Adhesión Marginal)



Media Benigno:	Desviación Estándar Benigno:	Media Maligno:	Desviación Estándar Maligno:
1.3468	0.9171	5.5858	3.1966

En la adhesión marginal tenemos datos dispersos en los tumores malignos, pero algo a resaltar es la concentración en los benignos, donde se ve claramente que la gran mayoría no sube de 1.

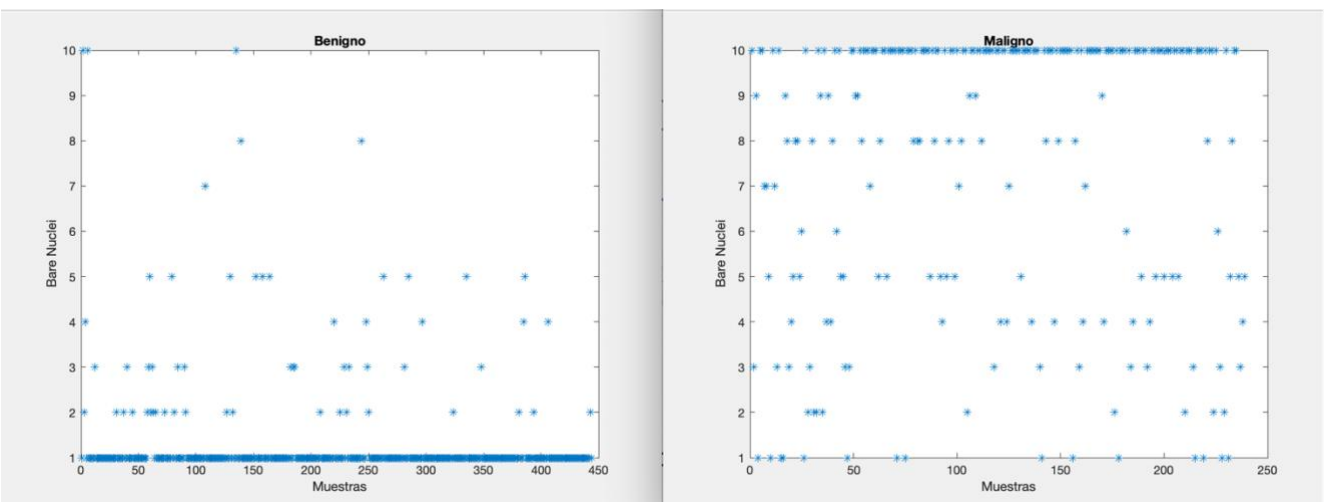
Single Epithelial Cell Size: 1-10 (Tamaño único de la célula epitelial)



Media Benigno:	Desviación Estándar Benigno:	Media Maligno:	Desviación Estándar Maligno:
2.1081	0.8771	5.3264	2.4431

Este dato puede ayudar a descartar un falso positivo (maligno), ya que es casi nula la probabilidad (en estos casos) que el tamaño de la célula epitelial sea menor a 2.

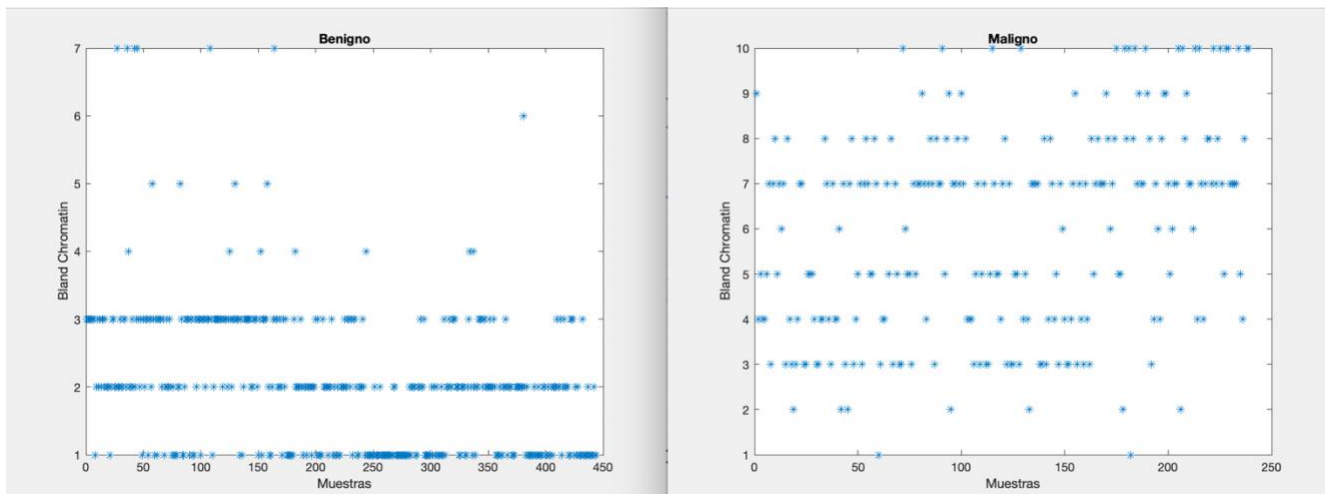
Bare Nuclei: 1-10 (Núcleos desnudos)



Media Benigno:	Desviacion Estandar Benigno:	Media Maligno:	Desviacion Estandar Maligno:
1.3468	1.1778	7.6276	3.1167

Podemos notar que el número núcleos desnudos(degenerados) es mucho mayor en los tumores malignos que en los benignos teniendo este una media de 7.6 con respecto a 1.3 de los benignos, lo que puede llegar a facilitar o evidenciar más rápido un diagnóstico.

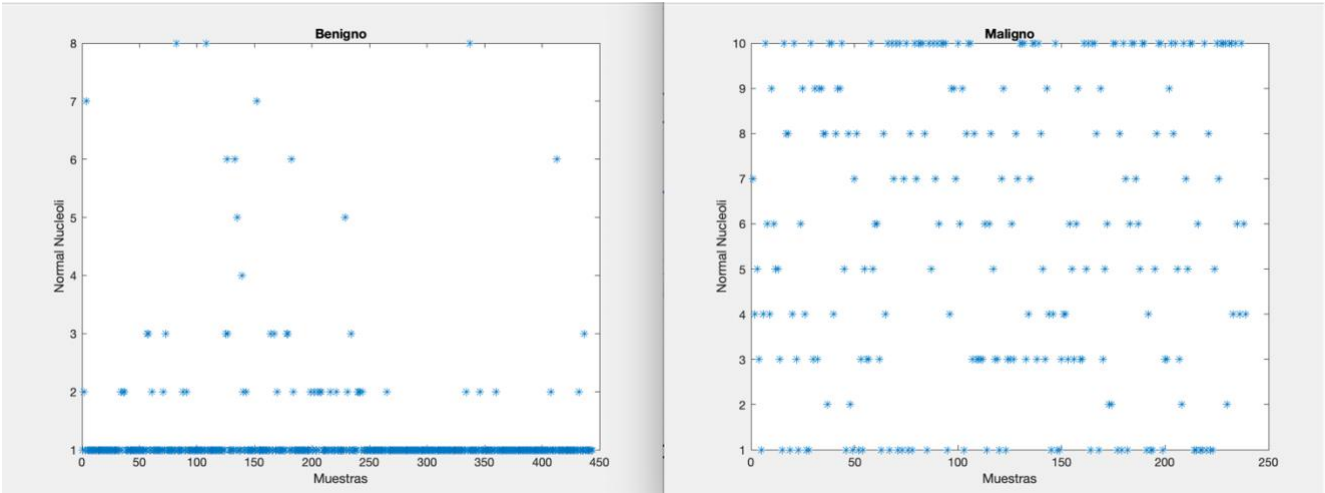
Bland Chromatin: 1-10 (Cromatina Blanda)Bland Chromatin: 1-10 (Cromatina Blanda)



Media Benigno:	Desviacion Estandar Benigno:	Media Maligno:	Desviacion Estandar Maligno:
2.0833	1.0623	5.9749	2.2824

Se puede observar fácilmente una diferencia entre la cromatina blanda de los tumores benignos y malignos, marcando los benignos una fuerte dispersión hacia el límite inferior con una media de 2.08 y una desviación de 1.06 mientras que la gráfica de los malignos nos muestra visiblemente una distribución mucho mayor con una tendencia hacia el límite superior.

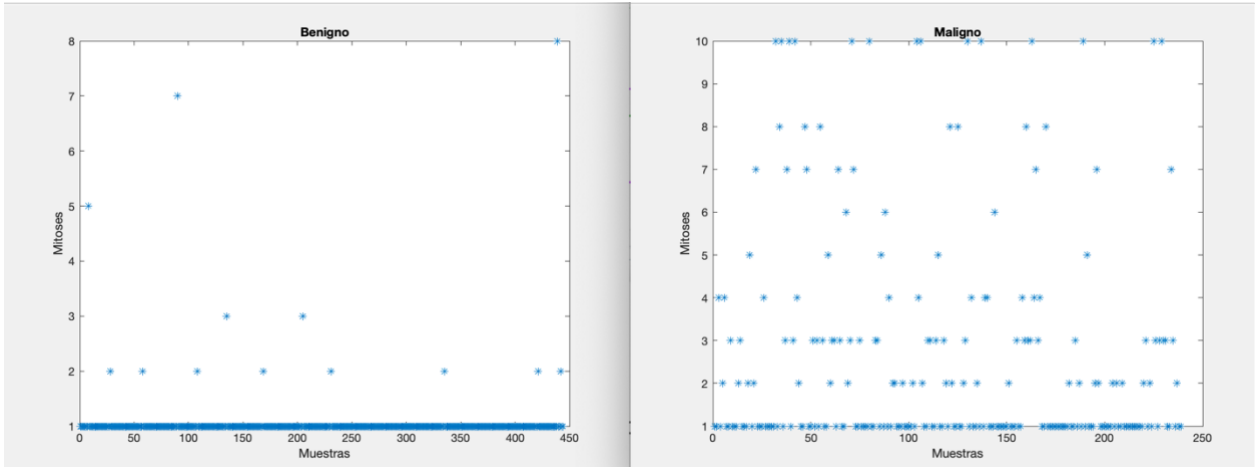
Normal Nucleoli: 1-10 (Núcleos Normales)



Media Benigno:	Desviación Estándar Benigno:	Media Maligno:	Desviación Estándar Maligno:
1.2613	0.9546	5.8577	3.3489

Se puede observar fácilmente que los núcleos normales benignos marcan una tendencia hacia uno con una media de 1.26 mientras que los malignos marcan una tendencia hacia 10 con una dispersión mucho mayor.

Mitoses: 1-10 (Mitosis)



Media Benigno:	Desviación Estándar Benigno:	Media Maligno:	Desviación Estándar Maligno:
1.0653	0.5097	2.6025	2.5645

En esta característica podemos notar que tanto los tumores benignos como malignos muestran una tendencia hacia el límite inferior, siendo sus medias de 1.06 y 2.60 respectivamente, pero, aún así podemos notar como los tumores malignos muestran una dispersión mayor de 2.56 con una tendencia hacia el límite superior .

Aplicando teorema de Bayes

• Conteo Normalizado Malignos

1	0.0182	0.0242	0.0182	0.1212	0.0121	0.0485	0.0182	0.1758	0.5273
2	0.0182	0.0364	0.0364	0.0848	0.0848	0.0485	0.0364	0.0364	0.1030
3	0.0545	0.1152	0.0788	0.1273	0.1758	0.0727	0.1576	0.1212	0.1455
4	0.0485	0.1333	0.1333	0.1152	0.1515	0.0545	0.1273	0.0667	0.0424
5	0.2061	0.0909	0.1333	0.0788	0.1394	0.0667	0.1212	0.0727	0.0182
6	0.0848	0.1030	0.1030	0.0848	0.1455	0.0303	0.0364	0.0727	0.0121
7	0.0667	0.1030	0.1152	0.0485	0.0303	0.0303	0.2606	0.0485	0.0303
8	0.1576	0.1152	0.1212	0.1152	0.0909	0.0727	0.1273	0.0606	0.0364
9	0.0606	0.0182	0.0364	0.0242	0.0182	0.0424	0.0485	0.0667	0.0061
10	0.2848	0.2606	0.2242	0.2000	0.1515	0.5333	0.0667	0.2788	0.0788

• Conteo Normalizado Benignos

1	0.3043	0.7960	0.7592	0.7993	0.1003	0.8428	0.3244	0.8428	0.9398
2	0.1003	0.0903	0.1104	0.0803	0.7726	0.0569	0.3378	0.0736	0.0201
3	0.1873	0.0669	0.0635	0.0635	0.0669	0.0368	0.2809	0.0334	0.0067
4	0.1773	0.0234	0.0368	0.0201	0.0167	0.0134	0.0100	0.0033	0.0033
5	0.1739	0.0033	0.0067	0.0134	0.0134	0.0234	0.0067	0.0100	0.0067
6	0.0368	0.0033	0.0067	0.0100	0.0067	0.0033	0.0067	0.0134	0.0033
7	0.0033	0.0033	0.0033	0.0033	0.0067	0.0067	0.0234	0.0100	0.0067
8	0.0100	0.0067	0.0067	0.0033	0.0067	0.0033	0.0033	0.0067	0.0067
9	0.0033	0.0033	0.0033	0.0033	0.0033	0.0033	0.0033	0.0033	0.0033
10	0.0033	0.0033	0.0033	0.0033	0.0067	0.0100	0.0033	0.0033	0.0033

Conclusiones

Probabilidades de que se den las siguientes clases:

Benigno: 0.6509

Maligno: 0.3491

Porcentaje de acierto dado el teorema:

Benigno: 100% (Falso negativo)

Maligno: 97.42% (Falso positivo)

En conclusión el teorema funciono correctamente con una tasa de acierto bastante alta y tomando en cuenta que es muy fiable no caer en un falso negativo (diagnosticar que no presenta problema un tumor cuando en realidad si), confiaría bastante en este análisis, siempre y cuando se auxilie un especialista con este análisis.