

PREGUNTA N 3

GALAXY ZOO: Reproducing galaxy morphologies via machine learning

1. Describa la funcion de costo del problema, en que consiste la muestra inicial y como esta clasificada.
2. Interprete los resultados del algoritmo mostrados en la figura 5 y explique como se genera la tabla 5.

Solucion:

1.
 - a) La funcion costo del problema es:

$$E = \sum_k (T_{NN}(w_{ij}, p_k) - T_{eye,k})^2,$$

Donde:

T_{NN} es la probabilidad de que la red neuronal pertenezca a un tipo de morfologia particular.

p_k Son los paramtetros de entrada hacia la red neuronal.

$T_{eye,k}$ Son las clasificaciones hechas por el “ojo humano”.

w_{ij} , Todos los nodos en una capa oculta entre las interconexiones y conexiones son nodos i y j, representados por este simbolo.

La funcion costo es usada en una red neuronal usada para el estudio de muchas capas, donde cada una consiste en un numero de nodos. La primera capa recibe los parametros de entrada. Y la ultima capa retorna la probabilidad de el objeto pertenece a una de las 3 clases morfologicas.

- b) la muestra inicial proviene de un catalogo de Galaxy Zoo, provista de un proyecto web que obtiene la clasificacion morfologica para millones de objetos y cada muestra esta clasificada por caractericticas especiales como se tiene a continuacion:

Table 1. First set of input parameters based on colours and profile fitting.

| Name | Description |
|------------------------|-------------------------------------|
| <i>dered_g-dered_r</i> | (<i>g</i> − <i>r</i>) colour |
| <i>dered_r-dered_i</i> | (<i>r</i> − <i>i</i>) colour |
| <i>deVAB_i</i> | de Vaucouleurs fit axial ratio |
| <i>expAB_i</i> | Exponential fit axial ratio |
| <i>lnLexp_i</i> | Exponential disc fit log likelihood |
| <i>lnLdeV_i</i> | de Vaucouleurs fit log likelihood |
| <i>lnLstar_i</i> | Star log likelihood |

Se tiene (g-r) y (r-i) colores derivados de los modelos, magnitudes que no tienen que ser k-correctos para el resto del marco.

El segundo conjunto de parametros son:

Table 2. Second set of input parameters based on adaptive moments.

| Name | Description |
|------------------------------|----------------------------|
| <i>petroR90_i/petroR50_i</i> | Concentration |
| <i>mRrCc_i</i> | Adaptive (+) shape measure |
| <i>aE_i</i> | Adaptive ellipticity |
| <i>mCr4_i</i> | Adaptive fourth moment |
| <i>texture_i</i> | Texture parameter |

q contienen nuevos conjuntos de formas y texturas, son usado para la concentracion. La concentracion es dada por la proporcion de radios contienen 90 y 50 porciento de flujo de petrosian. MrrCc es la segundo momento de intensidad en la CCD fila y columna usado en un esquema designado para obtener una senal optima senal de ruido de radio.

$$mRrCc = y^2 + x^2$$

donde x e y son imagenes coordinadas relativas al centro de l objeto en cuestion por ejemplo:

$$y^2 = \frac{\sum I(x,y)w(y,x)y^2}{\sum I(y,x)w(y,x)}$$

Los momentos se miden utilizando una función de peso Gauss radial, $w(y, x)$, adaptado de forma iterativa a la forma y tamaño del objeto. los componentes de elipticidad, mE1 y mE2, definidos en las ecuaciones (4) y (5) y los momentos de cuarto orden (ecuacion (6)) son de la manera:

$$mE1 = \frac{y^2 - x^2}{mRrCc} \dots(4), \quad mE2 = 2 \frac{yx}{mRrCc} \dots(5), \quad mCr4 = \frac{(x^2 + y^2)^2}{\sigma^2} \dots(6)$$

donde σ es el tamaño de los pesos Gaussianos. Se usa los componentes de elipticidad para definir la elipticidad adaptativa, aE, para ingresarla a la red neuronal

donde:

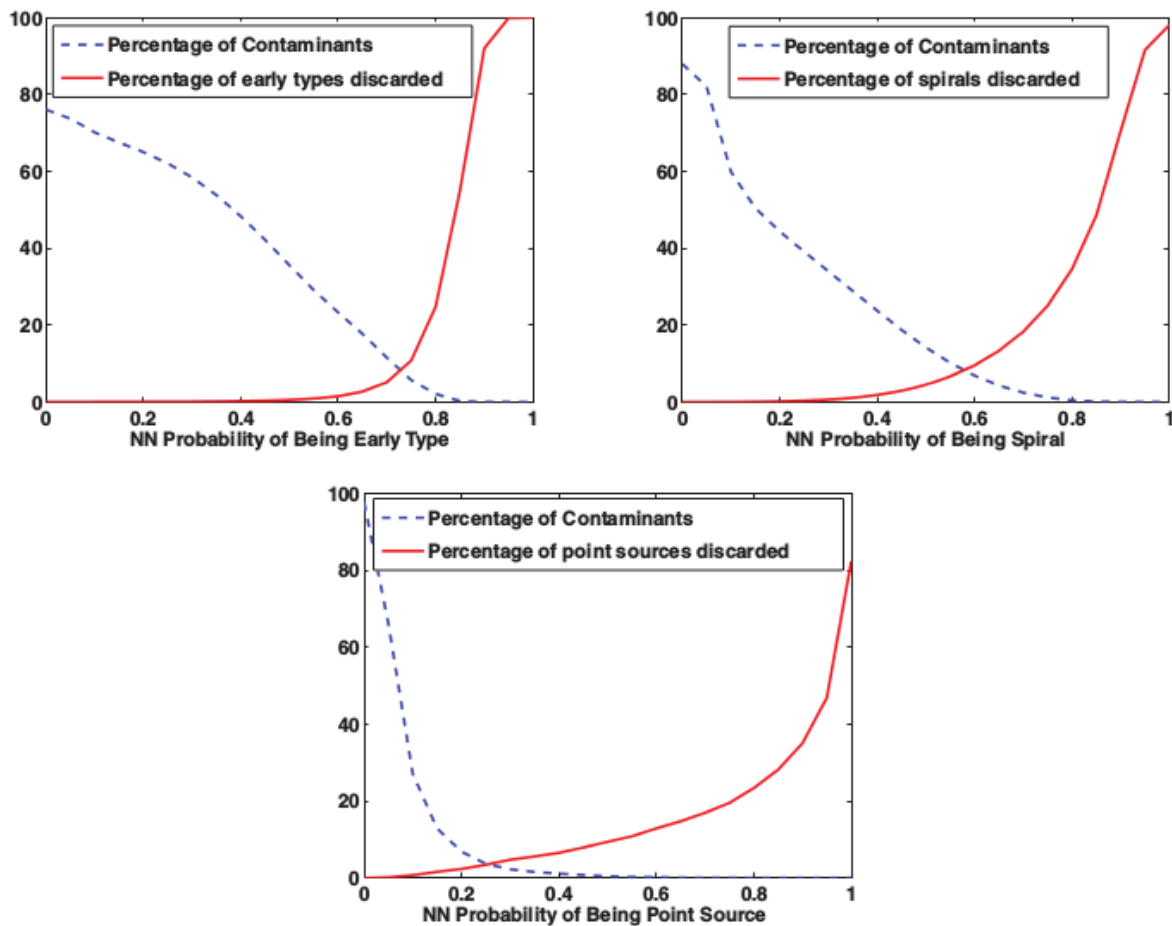
$$aE = 1 - \sqrt{\frac{1 - \sqrt{mE_1^2 + mE_2^2}}{1 + \sqrt{mE_1^2 + mE_2^2}}}$$
, el parametro final es esencialmente el radio del rango de fluctuaciones en la superficie del objeto.

c) Se clasifica en:

- I Galaxia tipo Spiral.
- II Galaxia tipo edad temprana.
- III Galaxia tipo punto de origen.

2.

a) Interprete los resultados del algoritmo mostrado de la figura 5.



Interpretacion:

Las figuras representan la probabilidad de red neuronal de una galaxia es un tipo temprano (arriba a la izquierda), en espiral (arriba a la derecha) y el punto de origen / artefacto (abajo) en comparación con el porcentaje de contaminantes, así como el porcentaje de Galaxy Zoo objetos en estas clases que se descartan. Estos resultados son obtenidos combinando el conjunto de datos de la tabla 1 y la tabla 2.

entonces las figuras muestran la grafica de la clase de morfologia versus el porcentaje de genuinidad del objeto en esa clase son descartados aplicando la probabilidad de contaminante(si pertenece o no)

b) Explique como se genera la tabla 5.

La tabla 5 se genera corriendo el algoritmo usando los 12 tipos de parametros de entrada con $r < 17$ (r es rango de contaminante).SE hace un Test y vemos que las galaxias brillantes en general tienen mejores clasificaciones comparadas en la tabla pero conforme vamos aumentando el radio desde $r < 17$ hasta $r < 17.77$ vemos que va cambiando el porcentaje de dichas cantidades. Las cuales se ven en las tablas del 5 al 9

Table 5. Summary of results for the entire sample when using input parameters specified in

Table 6. Summary of results for the gold sample when using input parameters specified in

Table 8. Summary of results for the gold sample when using input parameters specified in Tables 1 and 2

| | | Galaxy Zoo | | |
|---|-----------------------|--------------------------|----------------------|-------------------------------------|
| | | Early type (per cent) | Spiral (per cent) | Point source/artefact (per cent) |
| A | Early type | 97 | 0.2 | 1.3 |
| N | Spiral | 0.1 | 97 | 0.2 |
| N | Point source/artefact | 0.05 | 0.02 | 86 |

Table 9. Summary of results for the bright sample when using input parameters specified in Tables 1 and 2.

| | | Galaxy Zoo | | |
|---|-----------------------|--------------------------|----------------------|-------------------------------------|
| | | Early type (per cent) | Spiral (per cent) | Point source/artefact (per cent) |
| A | Early type | 94 | 0.1 | 0.3 |
| N | Spiral | 0.1 | 92 | 0.1 |
| N | Point source/artefact | 0.2 | 0.2 | 98 |

Una vez más el umbral de probabilidad requerido en la salida de la red neuronal para la clasificación se determina al requerir que el porcentaje de contaminantes para que sea igual al porcentaje de objetos genuinos en esa clase que se descartan en la aplicación del umbral. Este habilidad óptima probabilidad es muy similar para la muestra brillante a las mostradas en Fig. 5 pero ligeramente más altos en las tres clases cuando las galaxias brillantes se utilizan para el entrenamiento y se utiliza para clasificar todas las galaxias.

Comparando las Tablas 5 y 9, se puede observar que los resultados son ligeramente mejor cuando se realiza clasificaciones morfológicas en el brillante muestra con $r < 17$ en comparación con toda la muestra con $r < 17.77$, en ambos casos utilizando un conjunto de entrenamiento completo. Esto es de esperarse, ya que es más fácil de

distinguir entre los primeros tipos y espirales en un brillante muestra. Cuando se realiza la formación utilizando una formación incompleta establecen con $r < 17$ y todos los objetos con $r < 17,77$ clasifican, el red neuronal se las arregla para llevar a cabo estas clasificaciones de acuerdo a más del 90 % con los usuarios de Galaxy Zoo. Que son las personas que clasifican a simple vista (de aqui existe el parametro $T_{eye,k}$).