

DATABASE dis.data/dis.csv

INTRODUCCIÓN

La tiroides es una glándula en forma de mariposa ubicada en el cuello, justo arriba de la clavícula. Es una de las glándulas endocrinas que producen hormonas. Las hormonas tiroideas controlan el ritmo de muchas actividades del cuerpo. Estas incluyen la velocidad con la que se queman calorías y cuán rápido late el corazón. Todas estas actividades componen el metabolismo del cuerpo.

Los problemas tiroideos incluyen:

Bocio: Agrandamiento de la tiroides

Hipertiroidismo: Cuando la glándula tiroides produce más hormona tiroidea de lo que su cuerpo necesita

Hipotiroidismo: Cuando la glándula tiroides no produce suficiente hormona tiroidea

Cáncer de tiroides

Nódulos: Bultos en la tiroides

Tiroiditis: Hinchazón de la tiroides

Para el diagnóstico de enfermedades de la tiroides, los médicos usan su historia clínica, un examen físico y pruebas para la tiroides. A veces también utilizan una biopsia. El tratamiento depende del problema, pero puede incluir medicamentos, terapia con yodo radiactivo o cirugía de tiroides.

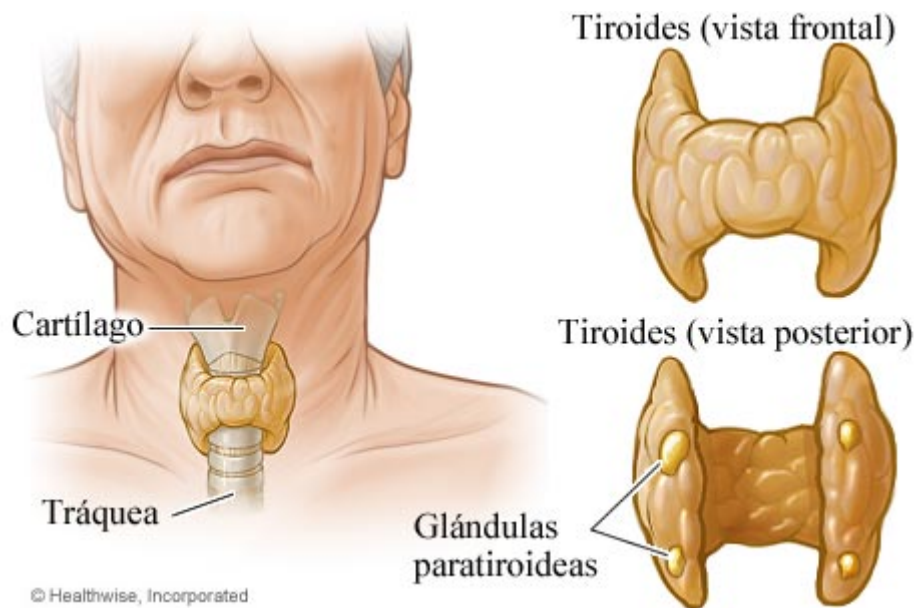
DESCRIPCIÓN DATA/ dis.csv

Registros de enfermedades de la tiroides suministrados por el Instituto Garavan y J. Ross Quinlan, Instituto de Nueva Gales del Sur, Sydney, Australia.;1987.

El problema es determinar si un paciente referido a la clínica es Hipotiroideo. Por lo tanto, se construyen tres clases: normal (no hipotiroideo), hiperfunción y funcionamiento subnormal. Porque el 92 por ciento de los pacientes no son hipotiroideos un buen clasificador debe ser significativamente mejor que el 92%.

Nota

Estos son los atributos que Quinlan's utilizó en el caso de estudio de su artículo "Simplificando los árboles de decisión" (International Journal of Man-Machine Studies) (1987) 221-234). Desafortunadamente estos datos difieren de la versión ya presente (donado por Ross Quinlan) Yo (Randolf Werner) no conozco más detalles sobre el conjunto de datos. Pero es difícil entrenar a las ANN de backpropagation con eso. El conjunto de datos se utiliza en dos informes técnicos (véase más arriba).



CONCLUSIÓN

Los datos de estudio en la database

Número de instancias: ann-train.data: 3772, ann-test.data: 3428

Número de clases: 3

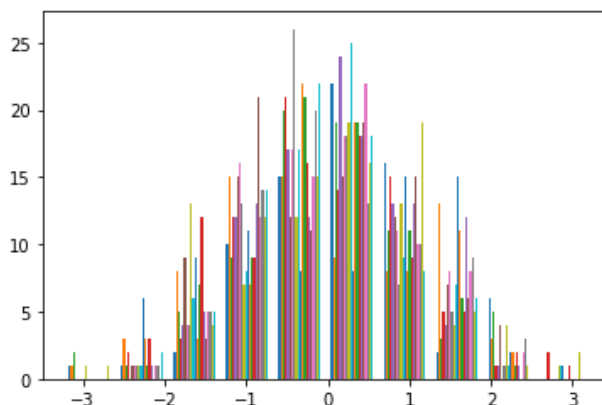
Número de atributos: 21 (15 atributos son binarios,
6 atributos son continuos)

Así decimos que el entrenamiento de datos que se realizó tomando los datos del transformacion del tipo array[]

En donde del database ha sido extraído El aprendizaje automático implica que una computadora sea entrenada usando un conjunto de datos dado y use esta capacitación para predecir las propiedades

```
In [349]: plt.hist(X_train)
```

```
Out[349]: (array([[ 1.,  1.,  2., 10., 15., 22., 16.,  2.,  6.,  0.],
 [ 1.,  3.,  8., 15., 15.,  9.,  8., 13.,  3.,  0.],
 [ 2.,  1.,  5.,  9., 20., 19., 11.,  3.,  5.,  0.],
 [ 0.,  2.,  3., 12., 21., 14., 15.,  5.,  1.,  2.],
 [ 0.,  0.,  4., 12., 17., 24., 13.,  4.,  1.,  0.],
 [ 0.,  1.,  9., 15., 12., 15., 12.,  7.,  4.,  0.],
 [ 0.,  1.,  4., 16., 17., 18., 11.,  8.,  0.,  0.],
 [ 0.,  1.,  4., 13., 26., 18.,  7.,  5.,  1.,  0.],
 [ 1.,  1., 13.,  7., 12., 19., 13.,  4.,  4.,  1.],
 [ 0.,  1.,  6.,  8., 17., 25.,  9.,  7.,  1.,  1.],
 [ 0.,  6.,  9., 11.,  8.,  8., 15., 15.,  2.,  1.],
 [ 0.,  3.,  3.,  7., 22., 19.,  8., 11.,  2.,  0.],
 [ 0.,  1.,  7.,  9., 21., 19., 11.,  6.,  1.,  0.],
 [ 0.,  3., 12.,  9., 16., 18.,  9.,  5.,  2.,  1.],
 [ 0.,  1.,  5., 13., 12., 18., 13., 12.,  1.,  0.],
 [ 0.,  0.,  3., 21., 11., 19., 15.,  6.,  0.,  0.],
 [ 0.,  1.,  5., 12., 15., 22., 10.,  8.,  2.,  0.],
 [ 0.,  1.,  5., 14., 20., 13., 10.,  9.,  3.,  0.],
 [ 1.,  0.,  4., 12., 15., 16., 19.,  5.,  1.,  2.],
 [ 0.,  2.,  5., 14., 22., 18.,  8.,  6.,  0.,  0.])),
array([-3.24126734, -2.59782985, -1.95439236, -1.31095487, -0.66751738,
       -0.02407989,  0.6193576 ,  1.2627951 ,  1.90623259,  2.54967008,
        3.19310757]),
<a list of 20 BarContainer objects>)
```



porcentaje medio en donde las personas con esta enfermedad
Dispersión se muestra en la figura ANTERIOR

1. Resumen de configuración de datos

Número de atributos: 21 (15 atributos son binarios,

6 atributos son continuos)

Número de clases: 3

Número de ejemplos de aprendizaje: 3772

Número de ejemplos de prueba: 3428

El conjunto de datos está disponible en un archivo ASCII

2. Descripción

El problema es determinar si un paciente derivado a la clínica es
hipotiroidismo

Por lo tanto, se construyen tres clases: normal (no hipotiroideo),

hiperfunción y funcionamiento subnormal. Porque el 92 por ciento de los pacientes no son hipertiroideos un buen clasificador debe ser significativamente mejor que 92%.

Nota

Estos son los datos que utilizó Quinlans en el caso de estudio de su artículo "Simplificación de árboles de decisión" (International Journal of Man-Estudios de máquinas (1987) 221-234)

HEADER DE LA TABLA

discordant, negative.	classes
age:	continuous.
sex:	M, F.
on thyroxine:	f, t.
query on thyroxine:	f, t.
on antithyroid medication:	f, t.
sick:	f, t.
pregnant:	f, t.
thyroid surgery:	f, t.
I131 treatment:	f, t.
query hypothyroid:	f, t.
query hyperthyroid:	f, t.
lithium:	f, t.
goitre:	f, t.
tumor:	f, t.
hypopituitary:	f, t.
psych:	f, t.
TSH measured:	f, t.
TSH:	continuous.
T3 measured:	f, t.
T3:	continuous.
TT4 measured:	f, t.
TT4:	continuous.
T4U measured:	f, t.
T4U:	continuous.
FTI measured:	f, t.
FTI:	continuous.
TBG measured:	f, t.
TBG:	continuous.
referral source:	WEST, STMW, SVHC, SVI, SVHD, other.

REFERENCIA

Saul Solorio Fernandez, SELECCIÓN DE VARIABLES PARA CLASIFICACIÓN NO SUPERVISADA UTILIZANDO UN ENFOQUE HÍBRIDO FILTER-WRAPPER , 2010