



Instituto Politécnico Nacional

ESCUELA SUPERIOR DE COMPUTO

LICENCIATURA EN CIENCIA DE DATOS

Aprendizaje de Máquina e Inteligencia Artificial

Grupo: 4AM1

Práctica III: Regresión Logística
Unidad 2: Aprendizaje Supervisado

FECHA DE ENTREGA: 7 DE ABRIL DE 2022

Profesor:

Joel Omar Gambino Juárez

Alumno:

De Luna Ocampo Yanina

Índice

1. Introducción	3
2. Pasos a seguir en esta práctica.	4
2.1. Tabla con los parámetros utilizados.	4
2.2. Exactitud promedio y exactitud obtenida con el modelo.	4
3. Conclusiones.	4

Aprendizaje de Máquina e Inteligencia Artificial

Práctica III: Regresión Logística

7 de abril de 2022

Resumen

En esta práctica pondremos a prueba los conocimientos que hemos revisado durante la clase de Aprendizaje Máquina e Inteligencia Artificial específicamente sobre el tema Regresión Lógica con la ayuda de scikit-learn. A continuación se dará una breve explicación de lo que buscamos resolver y con qué métodos lograremos lo pedido en las instrucciones de la práctica.

1. Introducción

Recordemos que para utilizar la regresión logística se necesita que la variable dependiente, o sea la variable objetivo, sea categórica, por ejemplo, predecir si un correo es spam o no, o ver si el tumor es maligno o benigno; como en esta práctica, entre muchos otros. Es semejante a un modelo de regresión lineal pero está adaptado para modelos en los cuales la variable dependiente es dicotómica, o sea 1 o 0. Es importante tener en cuenta que, aunque la regresión logística permite clasificar, se trata de un modelo de regresión que modela el logaritmo de la probabilidad de pertenecer a cada grupo. La asignación final se hace en función de las probabilidades predichas.

El dataset dado, tiene una lista de variables que describen el tamaño, radio, perímetro, textura y algunas otras características importantes del tumor. Los datos se refieren a los casos de pacientes con tumores mamarios y, como ya se ha dicho previamente, queremos predecir si el tumor es maligno o benigno con las características dadas. Las variables dadas dentro del dataset son las siguientes:

id	diagnosis	radius mean	texture mean	perimeter mean	area mean
smoothness mean	compactness mean	concavity _{mean}	concave points mean	symmetry mean	fractal dimension mean
radius se	texture se	perimeter se	area se	smoothness se	compactness se
concavity se	concave points se	symmetry se	fractal dimension se	radius worse	texture worst
perimeter worst	area worst	smoothness worst	compactness worst	concavity worst	concave points worst
symmetry worst	fractal dimension worst				

Tabla 1: Variables del dataset

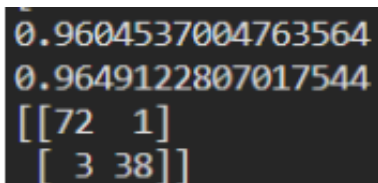
2. Pasos a seguir en esta práctica.

2.1. Tabla con los parámetros utilizados.

Pliegue	Modelo	Solver	Penalty	Iteraciones	Accuracy
1	Regresión Logística	lbfgs	none	10000	0.9671
2	Regresión Logística	lbfgs	none	10000	0.9473
3	Regresión Logística	lbfgs	none	10000	0.9668

Tabla 2: Modelo entrenado por pliegues.

2.2. Exactitud promedio y exactitud obtenida con el modelo.



```
0.9604537004763564
0.9649122807017544
[[72 1]]
[[ 3 38]]
```

3. Conclusiones.

Con esta práctica pusimos a prueba nuestros conocimientos con los temas vistos en un solo dataset. Entendimos y recordamos los objetivos que esta tiene, el primero es la probabilidad de ocurrencia de un suceso, la presencia o no de diversos factores y el valor o factor de los mismos y de segundo, determinar el modelo más ajustado que describa la relación entre la variable respuesta y un conjunto de variables regresoras. En esta práctica, el mejor valor que se obtuvo con el modelo fue de 0.96 de los datos de prueba. Sin embargo, podríamos tratar de aumentar la precisión mediante el uso de un algoritmo diferente que la regresión logística, o probar nuestro modelo con un conjunto diferente de variables.

Referencias

[Abu-Mostafa] Abu-Mostafa, Y. S. Learning From Data: A Short Course. Recuperado de: <https://work.caltech.edu/lectures.html>. Fecha de consulta: 20/03/2022.

[Scikit-Learn] Scikit-Learn. Linear model: Api reference. https://scikit-learn.org/stable/modules/classes.html#module-sklearn.linear_model. Fecha de consulta: 20/03/2022.