

Regresión Lineal Múltiple y Logística

García Prado, Sergio
sergio@garciparedes.me

Fernández Angulo, Óscar
oscar.fernandez.angulo@alumnos.uva.es

10 de mayo de 2017

Resumen

[TODO]

1. INTRODUCCIÓN

[TODO]

1.1. REGRESIÓN LINEAL MÚLTIPLE

[TODO]

1.2. REGRESIÓN LOGÍSTICA

[TODO]

2. EVALUACIÓN DE RESULTADOS A PARTIR DE DISTINTAS COTAS DE ERROR RELATIVO PARA REGRESIÓN LINEAL MÚLTIPLE

En esta sección se realiza un experimento para estudiar la evolución de la tasa de error conforme aumenta la cota de error relativo fijada para admitir una instancia como clasificada de manera correcta. El algoritmo utilizado es el de Regresión Lineal Múltiple mediante el aprendizaje por mínimos cuadrados descrito anteriormente.

El conjunto de datos que se ha utilizado para dichos experimentos es **Housing Data Set** [UCIa], que está formado por **506 instancias**. Dichas instancias contienen **13 atributos** de carácter numérico. La **clase de destino es de carácter numérico** (por lo que es una regresión y no una clasificación). En cuanto a la metodología experimental, se ha seguido una estrategia de **HoldOut** con particionamiento de los datos de manera que $\frac{2}{3}$ son utilizados en la fase de entrenamiento y $\frac{1}{3}$ en la de test.

Los resultados obtenidos tras los experimentos se muestran de manera tabular en la tabla 1 mientras que se puede apreciar la evolución conforme aumenta el rango de la cota de error de manera gráfica en la figura 1.

	Regresión — Housing Dataset			
	<i>Lineal 10 %</i>	<i>Lineal 15 %</i>	<i>Lineal 20 %</i>	<i>Lineal 25 %</i>
Error HoldOut	57.396 %	42.604 %	26.627 %	18.343 %

Tabla 1: Evolución de la tasa de error para la Regresión Lineal Múltiple sobre el conjunto de datos Housing conforme aumenta la cota máxima de error relativo

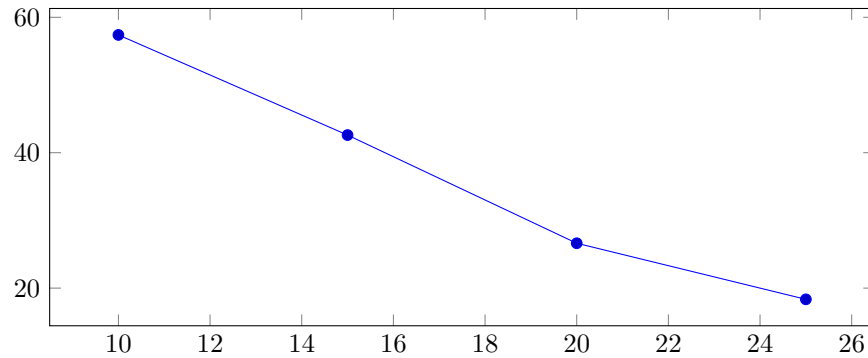


Figura 1: Evolución de la tasa de error para la Regresión Lineal Múltiple sobre el conjunto de datos Housing conforme aumenta la cota máxima de error relativo

3. COMPARACIÓN DE RESULTADOS ENTRE REGRESIÓN LOGÍSTICA Y REGRESIÓN LINEAL MÚLTIPLE

En esta sección se realiza un experimento para estudiar la tasa de error obtenida mediante la estrategia de *Regresión Logística* para después compararla con los resultados obtenidos mediante la estrategia de *Regresión Lineal*. Dicha labor es posible cuando la clase de destino es numérica pero se puede discretizar. (En este caso es de carácter entero y acotada por lo que su discretización es trivial). Puesto que no es de tipo binario es necesario seguir la estrategia de clasificación por pares tal y como se ha descrito anteriormente.

El conjunto de datos que se ha utilizado para dichos experimentos es **Wine Data Set** [UCIb], formado por **178 instancias**. Dichas instancias contienen **12 atributos** de carácter numérico. La **clase de destino es de carácter numérico**. Sin embargo, en este caso es de tipo entero, por lo que puede ser vista como una variable categórica. En cuanto a la metodología experimental, se ha seguido una extrategia de **HoldOut** con particionamiento de los datos de manera que $\frac{2}{3}$ son utilizados en la fase de entrenamiento y $\frac{1}{3}$ en la de test.

Los resultados obtenidos tras los experimentos se muestran de manera tabular en la tabla 2. Además, se muestran de manera gráfica en la figura 2.

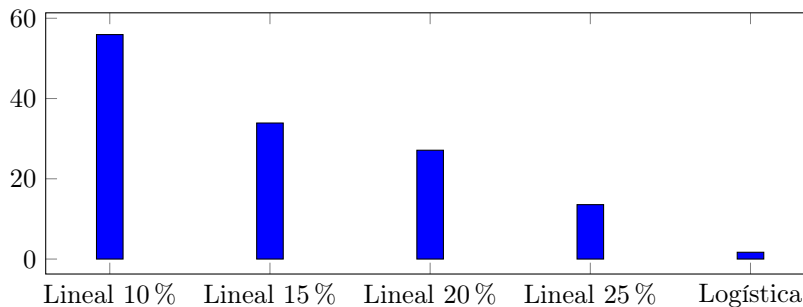


Figura 2: Resultados obtenidos a nivel de tasa de error mediante Regresión Lineal Múltiple y Regresión Logística sobre el conjunto de datos Wine

	Regresión — Wine Dataset				
	<i>Lineal 10 %</i>	<i>Lineal 15 %</i>	<i>Lineal 20 %</i>	<i>Lineal 25 %</i>	<i>Logística</i>
Error HoldOut	55.932 %	33.898 %	27.119 %	13.559 %	1.6949 %

Tabla 2: Resultados obtenidos a nivel de tasa de error mediante Regresión Lineal Múltiple y Regresión Logística sobre el conjunto de datos Wine

Tal y como se puede apreciar tras analizar los resultados, a través de la *Regresión Logística* se obtienen resultados mucho más óptimos. Sin embargo, para aplicar dicha técnica es necesario que la clase de destino esté discretizada. Además conlleva un mayor coste computacional derivado de la estrategia de clasificación por pares, que incrementa en un orden cuadrático los costes conforme aumentan el número de clases de destino.

REFERENCIAS

- [CCAG17] Teodoro Calonge Cano and Carlos Javier Alonso González. Técnicas de Aprendizaje Automático, 2016/17.
- [GP17] Sergio García Prado. Regresión lineal, múltiple y logística. <https://github.com/garciparedes/machine-learning-regression>, 2017.
- [too] Weka. <http://www.cs.waikato.ac.nz/ml/weka/>.
- [UCIa] UCI Machine Learning Repository. Housing Data Set. <http://archive.ics.uci.edu/ml/datasets/Housing>.
- [UCIb] UCI Machine Learning Repository. Wine Data Set. <http://archive.ics.uci.edu/ml/datasets/Wine>.