



ENFIN761

Business Intelligence para las Finanzas

AYUDANTÍA 6

Profesor: David Díaz S.

Ayudantes: Gabriel Cabrera G.¹

10 octubre 2019

Modelos Multinomial

El archivo `economic_stage.txt`, contiene variable macroeconómicas relevantes para explicar tres estado de la economía (*cluster-1*, *cluster-2* y *cluster-3*), a partir de lo visto en clases, responda las siguientes preguntas:

1. Incluya a su espacio de trabajo la base de datos, y realice un breve estadística descriptiva. Luego, separe las variables *features* y `economic_stage` (aquella que quiere predecir)
2. Genere una muestra de *Training* y *Testing*, está última debe ser igual al 50% del total de los datos y debe ser de manera aleatoria (*hint*: utilice `train_test_split` del módulo `model_selection` de la librería *skit-learn*).
3. Discuta las diferencias entre el algoritmo *One-versus-Rest* y *One-versus-All* en el contexto de regresión logística con y sin regularización.
4. Entrene un modelo mediante regresión logística utilizando el algoritmo *One-versus-Rest* con una regularización (penalización) del tipo *l1*. Realicelo usando `LogisticRegression` y luego con `sklearn.multiclass`. Presente la matriz de confusión (*Confusion Matrix*) tanto para la muestra de *Training* como *Testing* e interprete.
5. Entrene un modelo mediante regresión logística utilizando el algoritmo *One-versus-All* con una regularización (penalización) del tipo *l1*. Realicelo usando `sklearn.multiclass`. Presente la matriz de confusión (*Confusion Matrix*) tanto para la muestra de *Training* como *Testing* e interprete.
6. Entrene un modelo mediante regresión *Softmax* con una regularización (penalización) del tipo *l2*. Presente la matriz de confusión (*Confusion Matrix*) tanto para la muestra de *Training* como *Testing* e interprete.

Balanceo de Clases

1. Verifique si las clases están balanceado. En caso que no sea así, realice un sub-balanceo hacia la clase más pequeña. Luego, entrene un modelo mediante regresión logística utilizando el algoritmo *One-versus-Rest* con una regularización (penalización) del tipo *l1*. Realicelo usando `LogisticRegression`. Presente la matriz de confusión para ambas muestras e interprete.

¹✉:gcabrerag@fen.uchile.cl

2. Verifique si las clases están balanceado. En caso que no sea así, realice un sobre-balanceo (hacia la clase más grande). Luego, entrene un modelo mediante regresión logística utilizando el algoritmo *One-versus-Rest* con una regularización (penalización) del tipo *l1*. Realicelo usando `LogisticRegression`. Presente la matriz de confusión para ambas muestras e interprete.

Apéndice

La matriz de confusión (*confusion matrix*) para el caso binario se resume como:

Table 1: Posible resultados cuando se aplica un clasificador

	<i>Predicted class</i>		Total
	-	+	
-	True Neg. (TN)	False Pos. (FP)	N = TN + FP
+	False Neg. (FN)	True Pos. (TP)	P = FN + TP
Total	N* = TN + FN	P* = FP + TP	N* + P* = N + P

Donde se pueden obtener las siguientes medidas:

Table 2: Medidas para clasificaciones

Nombre	Definición	Sinónimo
Tasa Falso Positivo	FP/N	Error Tipo I, 1 - Specificity
Tasa Verdadero Positivo	TP/P	Sensitivity o Recall
Valores Positivos Predichos	TP/P^*	Precisión
Valores Negativos Predichos	TN/N^*	
F1	$2 \times \left\{ \frac{\frac{TP}{P} \times \frac{TP}{P^*}}{\frac{TP}{P} + \frac{TP}{P^*}} \right\}$	

El siguiente es un *output* directo desde python de una matriz de confusión (*Confusion Matrix*):

```
## array([[204,   5,   0],
##        [ 11,   7,   0],
##        [   1,   0,   6]])
```

Para obtener las métricas:

a. Recall (TP/P):

- i. $204/209 = 0.98$
- ii. $7/18 = 0.39$
- iii. $6/7 = 0.86$

b. Precisión (TP/P^*):

- i. $204/216 = 0.94$
- ii. $7/12 = 0.58$
- iii. $6/6 = 1$

El *F1-score* lo dejo como propuesto. Para verificar, El siguiente es un *output* directo desde python del *Classification Report*.

##		precision	recall	f1-score	support
##					
##	cluster-1	0.94	0.98	0.96	209
##	cluster-2	0.58	0.39	0.47	18
##	cluster-3	1.00	0.86	0.92	7
##					
##	micro avg	0.93	0.93	0.93	234
##	macro avg	0.84	0.74	0.78	234
##	weighted avg	0.92	0.93	0.92	234