

Algoritmos y técnicas de regresión

Conoce los modelos de regresión

El mapa de aplicaciones prácticas de ciencia de datos



Las técnicas de regresión

- Lineales
 - OLS – Regresión lineal
 - Stepwise para simplificar la regresión lineal
 - GLM – modelos lineales generalizados
 - Otras fórmulas de regresión (hay muchas...)
- NO lineales
 - CART – Árboles de clasificación y regresión
 - Random forest – Árboles aleatorios (generalización de CART)
 - KNN – k vecinos más cercanos
 - SVM – super vectors machine
 - NN – redes neuronales
 - Algoritmos de bagging

Son los que has visto para clasificación adaptados a regresión

- En esta lección vamos a trabajar con la implementación de los algoritmos que has visto en regresión
- Te desvelaré algo interesante para la regresión – el stepwise

Optimizando la regresión

El stepwise para optimizar la regresión

Stepwise es sinónimo de escoger el mejor modelo de regresión lineal

- Stepwise consigue ajustar un modelo que sea el compromiso entre:
 - Minimizar el error
 - Las menos variables posibles
- En el fondo consigue seleccionarte el conjunto de variables de entrada más relevantes (estadísticamente claro)
- Para ello utiliza el AIC, un índice que penaliza la complejidad del modelo y premia minimizar el error

Un ejemplo de stepwise

- Además puedes calcular el rango de importancia de las variables de entrada
- Mira:

Call:
lm(formula = Employed ~ ., data = datos)

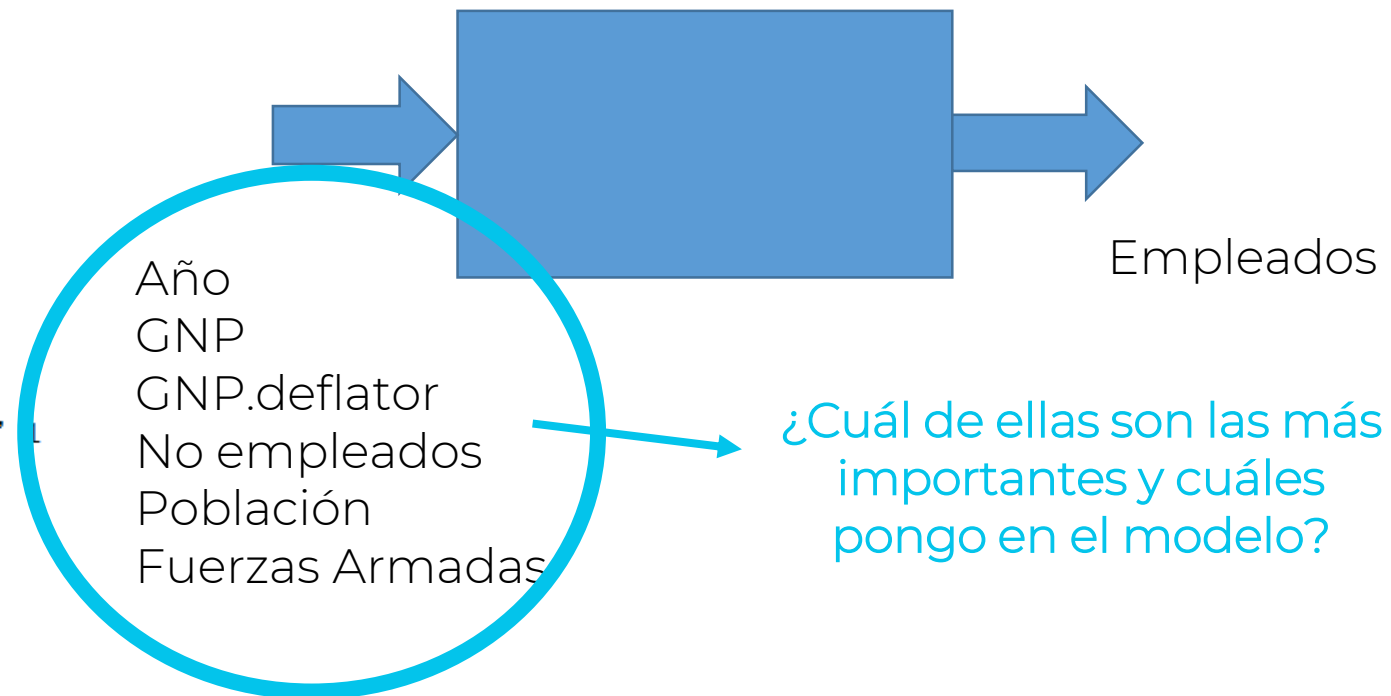
Residuals:
Min 1Q Median 3Q Max
-0.41011 -0.15767 -0.02816 0.10155 0.45539

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-3.482e+03	8.904e+02	-3.911	0.003560	**
GNP.deflator	1.506e-02	8.492e-02	0.177	0.863141	
GNP	-3.582e-02	3.349e-02	-1.070	0.312681	
Unemployed	-2.020e-02	4.884e-03	-4.136	0.002535	**
Armed.Forces	-1.033e-02	2.143e-03	-4.822	0.000944	***
Population	-5.110e-02	2.261e-01	-0.226	0.826212	
Year	1.829e+00	4.555e-01	4.016	0.003037	**

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3049 on 9 degrees of freedom
Multiple R-squared: 0.9955, Adjusted R-squared: 0.9925
F-statistic: 330.3 on 6 and 9 DF, p-value: 4.984e-10



Un ejemplo de stepwise

- Además puedes calcular el rango de importancia de las variables de entrada
- Mira:

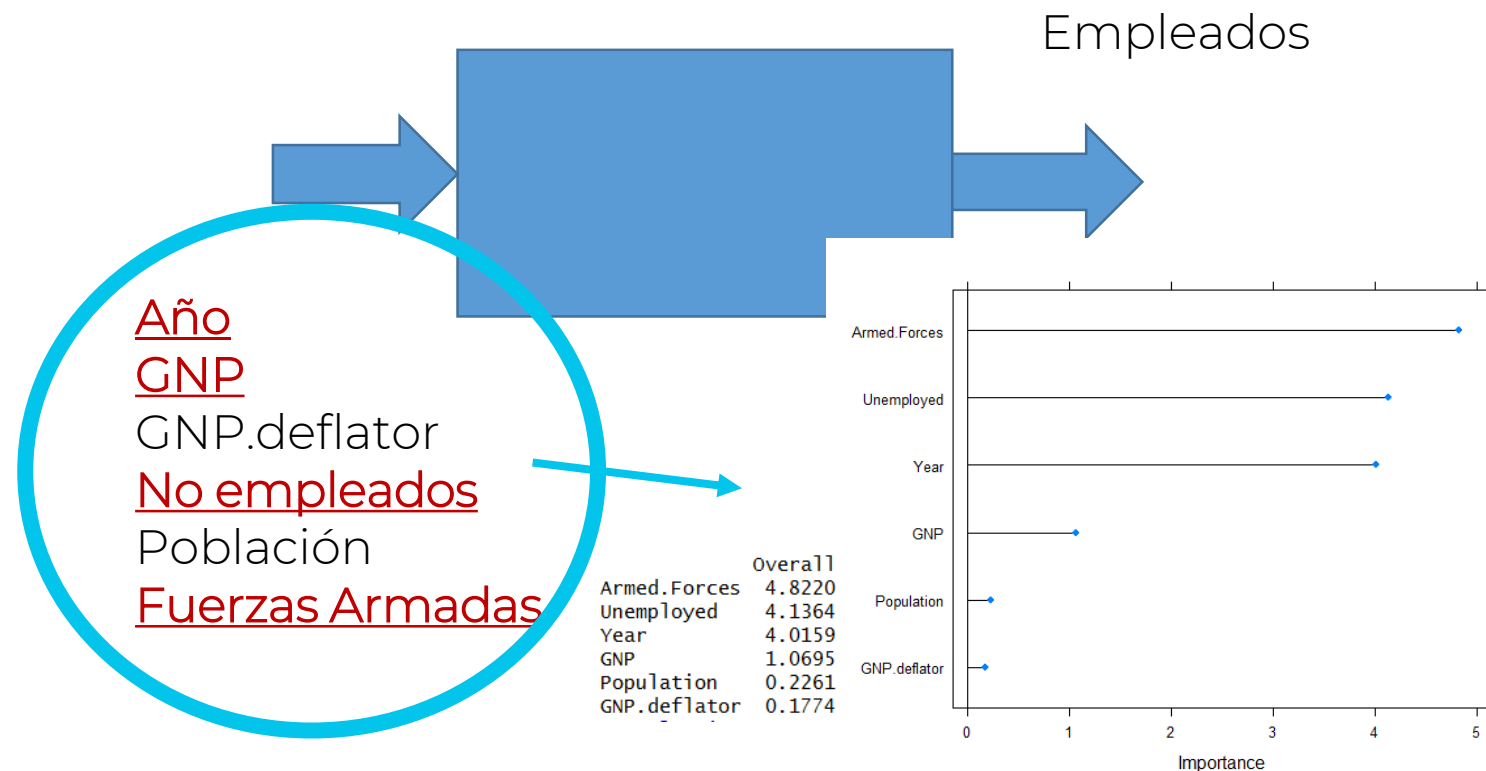
```
Call:
lm(formula = Employed ~ GNP + Unemployed + Armed.Forces + Year,
    data = datos)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.42165	-0.12457	-0.02416	0.08369	0.45268

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-3.599e+03	7.406e+02	-4.859	0.000503	***
GNP	-4.019e-02	1.647e-02	-2.440	0.032833	*
Unemployed	-2.088e-02	2.900e-03	-7.202	1.75e-05	***
Armed.Forces	-1.015e-02	1.837e-03	-5.522	0.000180	***
Year	1.887e+00	3.828e-01	4.931	0.000449	***



Tú turno

Un pequeño ejercicio para empezar

Tú turno

- Descarga la hoja de trabajo que te he preparado
- Mira los vídeos de la zona tech zen y con la ayuda de las recetas de código responde a las preguntas para asentar cómo implementar estas técnicas