



Afi

Escuela
de Finanzas

Regresión Avanzada

Máster en Data Science y Big Data en Finanzas

Javier Nogales – PhD Matemáticas
Profesor Titular, Estadística UC3M

www.est.uc3m.es/nogales @fjnogales

Ejercicio GLM + Regresión Avanzada

Usar herramientas vistas en clase para extraer conclusiones sobre datos Covid

- Existen muchas fuentes de datos Covid abiertos:

▶ Our World in Data

▶ Johns Hopkins University

▶ European Centre for Disease Prevention and Control

▶ Datadista

▶ ISCI

▶ Escovid19data

▶ Financial Times

- Usar alguna de estas fuentes (u otras) y desarrollar modelos GLM y de Regresión Avanzada para entender mejor las relaciones entre variables
- Posibles variables target: número de fallecidos, exceso de mortalidad, tasa de letalidad, número de hospitalizados/UCI, casos diagnosticados, IA14, tasa de cambio (dato hoy - dato ayer), etc.
- Se puede hacer a nivel diario, mensual, etc. Se puede hacer a nivel país, provincia, etc.

GLM: 40 %

- **Objetivo:** desarrollar modelos GLM (que no sean LM) que expliquen una variable target en función de predictores
- Proporcionar *interpretación* y *conclusiones*
- Evaluar capacidad predictiva (dividiendo muestra en train y test), considerando intervalos de predicción
- Otros posibles objetivos: predecir casos primera ola en función de datos segunda y tercera ola, efecto de confinamientos, análisis de decalajes entre variables (primero casos, luego hospitalizados, luego UCI, luego fallecidos), etc.
- Se aconseja usar más de una fuente de datos para conseguir mejores predictores

GLM: 40 %

- Usar preferiblemente datos a partir de 1 de julio 2020, por fiabilidad, salvo que se quiera predecir primera ola
- Se aconseja suavizar las variables (p.e. tomando medias móviles, o IA14, etc.), sobre todo si se usan datos diarios
- Si se quiere explicar una tasa (p.e. fallecidos entre población), mejor usar fallecidos y luego offset en modelo Poisson: $\text{offset} = \log(\text{population})$
- Se valora originalidad en el planteamiento del problema (variable target y predictores)
- Se permite coger inspiración en otros análisis, citando siempre la fuente
- Subir a campus virtual: datos, notebook (Rmd o Jupyter) y versión compilada (pdf o html)

Advanced Regression: 60 %

- **Objetivo:** desarrollar herramientas de regresión avanzada (no GLM) para predecir una variable target en función de predictores.
Se aconseja que la variable target sea distinta a la de la primera parte del ejercicio.
- En esta parte es obligatorio fusionar dos bases de datos distintas para conseguir más predictores (aunque no influyan) y trabajar en *alta dimensión*
- Usar las herramientas vistas en clase para obtener un modelo final que explique la variable target (usando toda la muestra) y otro que obtenga predicciones de dicha variable (para muestra test) junto con intervalos de predicción
- Se valora originalidad: p.e. considerar interacciones entre variables, reducción de la dimensionalidad del problema, extracción de features, etc.
- Se permite coger inspiración en otros análisis, citando siempre la fuente
- Subir a campus virtual: datos, notebook (Rmd o Jupyter) y versión compilada (pdf o html)

Evaluación

- GLM preparación datos: 0.5 puntos
- GLM desarrollo e interpretación de modelos: 2 puntos
- GLM predicción: 0.5 puntos
- GLM originalidad y report: 1 punto
- Regresión Avanzada preparación datos: 1 punto
- Regresión Avanzada parte explicativa: 2 puntos
- Regresión Avanzada parte predictiva: 2 puntos
- Regresión Avanzada originalidad y report: 1 punto



Afi

Escuela
de Finanzas

© 2015 Afi Escuela de Finanzas. Todos los derechos reservados.