



ENFIN761

Business Intelligence para las Finanzas

AYUDANTÍA 5

Profesor: *David Díaz S.*

Ayudantes: *Gabriel Cabrera G.*¹

02 octubre 2019

Regresión Polinomial

1. Genere una variable X e y que contengan 100 observaciones con las siguientes características:

- a. $X = 6 \cdot \text{Aleatorio}_1 - 3$
- b. $y = 2 + X + 0.35 \cdot X^2 + 0.20 \cdot X^3 + \text{Aleatorio}_2$

El parámetro Aleatorio_1 se refiere a la función `np.random.rand()` y Aleatorio_2 a la función `np.random.randn()`. Utilice una semilla igual a 42.

2. Vizualice los datos generados en (1).
3. Estime una regresión lineal con constante, luego:

$$y_t = \theta_0 + \theta_1 x_t + \varepsilon_t$$

- a. Calcule las métricas RMSE (*Root Mean Square Error*) y R^2 .
 - b. Realice la predicción utilizando los parámetros $\hat{\theta}$ encontrados en (3.a).
 - c. Agregue la predicción al gráfico generado en (2).
4. Estime una regresión polinomial con grados igual a 2, 3 y 30. Para cada estimación considere:
 - a. Almacene en una lista o *DataFrame* las métricas:
 - i. RMSE (*Root Mean Square Error*).
 - ii. R^2 .
 - b. Grafique los datos generados y la predicción por cada grado del polinomio.
 5. A partir de lo encontrado en (4), comparé su resultados con la regresión lineal en términos visuales, coeficientes y métricas de *performance*.
 6. Genere una función que permita visualizar la *learning curves* de cada estimación realizada hasta ahora. Discuta los resultados usando los conceptos de *Underfitting* y *Overfitting*.

¹✉:gcabrera@fen.uchile.cl

Modelos Regularizados

En el archivo `returns_short_interest_data.xlsx` se encuentran las variables necesarias para construir los 14 predictores² utilizados por David Rapach, Matthew Ringgenberg y Guofu Zhou en su trabajo empírico, *Short Interest and Aggregate Stock Returns (2016)* (adjunto). A partir de estos predictores, responda las siguientes preguntas.

1. Construya los 14 predictores.
2. Genere la estadística descriptiva de cada predictor. Debe contener, promedio (*mean*), mediana (*median*), 1th y 99th percentil y desviación estandar. respete el orden y redonde al segundo decimal.
3. Genere la matriz de correlaciones usando un DataFrame y luego presentelo visualmente.
4. Utilizando la variable `CRSP_SPvw` (retorno del S&P 500 incluyendo dividendos calculado por el CRSP), genere:
 - a. El *equity premium*.
5. Considerando el siguiente modelo univariado:

$$r_t = \theta_0 + \theta_1 x_t + \varepsilon_t$$

Donde r_t es el *equity premium* en logaritmo para el mes t y x_t el predictor. Estime mediante MCO la regresión utilizando los 14 predictores. En cada regresión almacene el coeficiente, estadístico t , p -value, R^2 y RMSE (*Root Mean Squared Error*).

6. Utilizando como fecha de corte, Enero del 2004, genere una muestra de *Training* y *Testing*. A partir de la muestra de *Training*, estime mediante Regresión lineal, Ridge, Lasso y Elastic Net (modelos regularizados), la predicción del *equity premium* usando todos los predictores (14). Calcule el MAE (*Mean Absolute Error*) y RMSE (*Root Mean Square Error*):

$$r_t = \theta_0 + \sum_{j=1}^k \theta_j x_{jt} \quad \text{s.a.} \quad \underbrace{\alpha \frac{1}{2} \sum_{j=1}^k \theta_j^2}_{\text{Ridge}} + \underbrace{\alpha \sum_{j=1}^k |\theta_j|}_{\text{Lasso}} \\ \text{Elastic Net } (r)$$

Tanto la regresión Ridge y Lasso, deben tener un λ igual a 0.2.

7. Verifique como se comporta cada modelo en la muestra de *Testing* utilizando las medidas MAE (*Mean Absolute Error*) y RMSE (*Root Mean Square Error*).
8. Estime el modelo presentado en (4) usando una regresión Ridge, con una $\lambda = 0.5, 5, 10, 20, 50, 100, 150, 200, 1000$ y 5000 . Esta vez muestre el MSE (*Mean Square Error*) por cada λ .

²Los predictores provienen del trabajo de Ivo Welch y Amit Goyal, *A Comprehensive Look at The Empirical Performance of Equity Premium Prediction (2008)*.