

Tarea: Modelos de Regresión Aplicados a Fondos de Inversión y Factores de Riesgo

Aplicaciones Profesionales en Ciencia de Datos

Objetivo General de la Tarea

En esta entrega, se pretende que los estudiantes apliquen métodos de regresión para predecir los rendimientos de distintos fondos de inversión utilizando como variables explicativas factores de riesgo que suelen tener buen poder predictivo. Los métodos de regresión que deberán emplear son:

- Regresión Lineal Ordinaria (OLS).
- Lasso.
- Ridge.
- Elastic Net.

Adicionalmente, se realizará un análisis de desempeño de los modelos a través de la división de los datos en conjuntos de entrenamiento y prueba (*training* y *test*), y se evaluará una estrategia de inversión simple basada en la predicción de los retornos.

1 Descripción de los Datos

El archivo Excel proporcionado contiene dos hojas:

- **Hoja 1:** Rendimientos mensuales de varios fondos de inversión. Estos rendimientos están expresados en términos decimales (por ejemplo, 0.02 significa un 2% de rendimiento).
- **Hoja 2:** Factores de riesgo. Cada columna corresponde a uno de los factores. Estos factores suelen encontrarse multiplicados por 100 en las bases de datos públicas, por lo que en este caso es **importante notar que los factores están multiplicados por 100**, mientras que los rendimientos de los fondos no lo están.

1.1 ¿Qué son los factores de riesgo?

En finanzas, uno de los modelos más usados para explicar los rendimientos de activos (acciones, fondos de inversión, etc.) es el de Fama-French, que originalmente incluía tres factores:

1. **MKT-RF** (Exceso de rendimiento del mercado sobre la tasa libre de riesgo).
2. **SMB** (Small Minus Big): Factor que captura la prima de tamaño.
3. **HML** (High Minus Low): Factor que captura la prima de valor.

Posteriormente, se ampliaron a cinco factores:

1. **RMW** (Robust Minus Weak): Factor de rentabilidad (empresas con alta rentabilidad menos empresas con baja rentabilidad).
2. **CMA** (Conservative Minus Aggressive): Factor de inversión (empresas conservadoras menos empresas agresivas).

Además, se agrega un factor adicional de **Momentum** (UMD: Up Minus Down), que captura la tendencia de los activos a continuar con su dirección de mercado reciente.

Estos factores se utilizan para explicar (o descomponer) el rendimiento de un activo. Por ejemplo, si un fondo invierte mayoritariamente en acciones de empresas pequeñas con alta rentabilidad, podríamos esperar que sus rendimientos se expliquen en buena parte por los factores SMB y RMW.

1.2 Importancia de Regresar Rendimientos de Fondos en Función de Factores

En el ámbito de las inversiones, es común medir hasta qué punto los rendimientos de un fondo son atribuibles a ciertas exposiciones a factores de mercado. De esta manera, podemos:

- Identificar si el rendimiento proviene de *alpha* (habilidad del gestor) o simplemente de la exposición a factores de riesgo comunes en el mercado.
- Comparar distintos fondos en función de sus *betas* (sensibilidades) a cada factor y su capacidad de generar rendimientos *ajustados al riesgo*.

2 Actividades a Realizar

A continuación se describe paso a paso lo que deben llevar a cabo.

2.1 Paso 1: Análisis de los Datos

1. **Carga de Datos:** Importar el archivo Excel. Asegúrense de identificar correctamente las columnas de la hoja de los rendimientos de los fondos (*fund returns*) y la hoja de los factores Fama-French + Momentum.
2. **Transformaciones Iniciales:**
 - Verifiquen que los rendimientos de los fondos y los factores de riesgo se encuentran en la misma escala. De lo contrario, realicen las transformaciones necesarias para asegurarse de ello.
 - Verifique si es necesario o no estandarizar las variables explicativas para su posterior utilización en los métodos de regresión.

2.2 Paso 2: Regresión con Distintos Modelos

Para cada fondo:

1. Entrenar un modelo de **Regresión Lineal Ordinaria (OLS)**:

$$\text{Return.Fondo}_t = \alpha + \beta_1(\text{Factor1}_t) + \beta_2(\text{Factor2}_t) + \dots + \epsilon_t$$

donde $\text{Factor1}, \text{Factor2}, \dots$ son MKT-RF, SMB, HML, RMW, CMA y Momentum.

2. Entrenar un modelo de regresión **Lasso**.

3. Entrenar un modelo de regresión **Ridge**.
4. Entrenar un modelo de regresión **Elastic Net**.
5. **Comparar los coeficientes** resultantes de cada modelo. Observen qué tan distintos son y, en particular, si ciertos factores parecen más relevantes (coeficientes más grandes) bajo un método de regularización que bajo otro.

2.3 Paso 3: División de Datos (Training y Test)

1. Dividan la muestra temporalmente (o aleatoriamente, de acuerdo con sus criterios) en dos:
 - Conjunto de entrenamiento (*training*): aproximadamente el 70-80% de los datos.
 - Conjunto de prueba (*test*): el restante 20-30%.
2. Ajusten los cuatro modelos (OLS, Lasso, Ridge, Elastic Net) en el conjunto de entrenamiento.
3. Estimen el R^2 de cada modelo tanto en *training* como en *test*. Discutan las diferencias encontradas en términos de sobreajuste (*overfitting*) o capacidad de generalización.

2.4 Paso 4: Análisis de Inversión

1. Utilizando las predicciones de cada modelo en el **conjunto de entrenamiento**, ordenen los fondos de acuerdo con su rendimiento esperado (predicho) desde el más alto hasta el más bajo.
2. Identifiquen los fondos que pertenecen al **top decile** (decil superior) y el **bottom decile** (decil inferior) de las predicciones.
3. Calcular los rendimientos reales en el **conjunto de prueba** de los fondos que quedaron en el top decile y en el bottom decile. Comparen esos rendimientos promedios y discutan los resultados.
 - ¿Los fondos que se esperaba tuvieran mejor rendimiento (según los modelos) en el período de entrenamiento efectivamente obtuvieron mejores resultados en el período de prueba?
 - ¿Hay diferencias marcadas entre los métodos (OLS, Lasso, Ridge, Elastic Net) al seleccionar los fondos?

3 Indicaciones Generales

- La tarea se puede realizar en grupos de **máximo 2 estudiantes**.
- **Entregable:** Un **único fichero PDF** (máximo 10 páginas) que incluya el análisis, las comparaciones de los modelos y las reflexiones sobre los resultados.
- **Utilización de Inteligencia Artificial:** Se asume que se utilizará algún tipo de IA (ChatGPT, Claude, DeepSeek, etc.) como apoyo para esta tarea. A raíz de ello, se espera que la calidad de la entrega sea superior a la que se obtendría sin el uso de estas herramientas. Por consiguiente, la corrección se realizará teniendo en cuenta la disponibilidad de este tipo de asistentes.
- **Plazo de entrega:** 23:59 h del día 20/02/2025 a través de **Canvas**.