Transformaciones, regresión y “destransformación” de la serie

En esta sección explicamos con todo detalle qué está pasando en el módulo `py103.py`, especialmente en el caso de la transformación por diferencias (`diff_close`), y aclaramos exactamente qué se está “destransformando” cuando dibujamos la serie en el espacio original de precios.

1. Notación básica

- Sea $t \in \{0, 1, \ldots, n-1\}$ el índice de tiempo discreto (días).
- Sea P_t el precio de cierre (`close`) de NVDA en el día t.
- Sea $\mathcal{D} = \{(t, P_t)\}_{t=0}^{n-1}$ nuestra serie temporal.
- Sea $g(\cdot)$ una transformación escalar aplicada a P_t.
- Sea $y_t = g(P_\bullet)$ el target transformado que realmente usamos en la regresión.

En el código, usamos varios g distintos:

$$\begin{aligned}
\texttt{close}: \quad & y_t = P_t, \\
\texttt{log_close}: \quad & y_t = \log P_t, \\
\texttt{diff_close}: \quad & y_t = P_t - P_{t-1}, \\
\texttt{simple_return}: \quad & y_t = \frac{P_t - P_{t-1}}{P_{t-1}}, \\
\texttt{log_return}: \quad & y_t = \log P_t - \log P_{t-1},
\end{aligned}$$

etc. (para algunos t iniciales, la serie se recorta o se rellena de forma conveniente; en el código lo manejamos con máscaras y `dropna`).

2. Construcción del problema de regresión

El modelo de regresión que usamos es un polinomio en el tiempo t, regularizado con Elastic Net.

2.1. Variable explicativa: índice de tiempo

Definimos el índice de tiempo

$$x_t = t \in \mathbb{R}, \qquad t = 0, 1, \ldots, n-1.$$

En notación matricial,

$$X = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ n-1 \end{bmatrix} \in \mathbb{R}^{n\times 1}.$$

Si usamos un polinomio de grado d en t, internamente aplicamos `PolynomialFeatures`:

$$\phi_d(t) = \big(t, t^2, \ldots, t^d\big),$$

y el diseño completo es

$$Z = \big[\phi_d(t_0)^\top;\ \phi_d(t_1)^\top;\ \ldots;\ \phi_d(t_{n-1})^\top\big] \in \mathbb{R}^{n\times d}.$$

2.2. Variable respuesta: transformación de la serie

Dado un esquema de transformación g, construimos

$$y_t = g(P_\bullet), \qquad t \in \mathcal{I}_g,$$

donde $\mathcal{I}_g \subseteq \{0, \ldots, n-1\}$ es el conjunto de instantes para los que la transformación está bien definida (por ejemplo, para `diff_close` empezamos en $t = 1$ y ajustamos índices).

En forma vectorial escribimos

$$y = (y_t)_{t\in\mathcal{I}_g} \in \mathbb{R}^m,$$

donde $m = |\mathcal{I}_g|$.

2.3. Separación en train / val / test

Ordenando por tiempo, dividimos y y X en tres segmentos contiguos:

$$\begin{aligned} \text{train: } & t = 0, \ldots, n_{\text{train}} - 1, \\ \text{val: } & t = n_{\text{train}}, \ldots, n_{\text{train}} + n_{\text{val}} - 1, \\ \text{test: } & t = n_{\text{train}} + n_{\text{val}}, \ldots, n - 1. \end{aligned}$$

En el código usamos proporciones

$$\texttt{train\_frac} = 0{,}6, \quad \texttt{val\_frac} = 0{,}2, \quad \text{resto} = 0{,}2 \text{ para test.}$$

3. Modelo de regresión en el espacio transformado

El modelo es siempre una regresión lineal sobre las features polinomiales, regularizada con Elastic Net. En notación estadística:

$$y_t = f_\theta(t) + \varepsilon_t, \qquad t \in \mathcal{I}_g,$$

donde

$$f_\theta(t) = \beta_0 + \beta_1 t + \beta_2 t^2 + \cdots + \beta_d t^d$$

(es decir, el modelo es lineal en los parámetros β, pero no en t), y ε_t es el error.

El ajuste se hace resolviendo

$$\hat{\theta} = \arg\min_\theta \left\{ \sum_{t \in \text{train}} \big(y_t - f_\theta(t)\big)^2 + \lambda \left[(1-\alpha)\|\theta\|_2^2 + \alpha\|\theta\|_1\right] \right\},$$

donde:

- $\lambda =$ `alpha` en el código.
- $\alpha =$ `l1_ratio` controla la mezcla Lasso/Ridge.

El modelo ajustado produce, para cada t,

$$\hat{y}_t = f_{\hat{\theta}}(t).$$

4. Caso especial: transformación de diferencias

Para `diff_close`, definimos

$$D_t = P_t - P_{t-1}, \qquad t = 1, \ldots, n-1.$$

En la práctica:

- Acomodamos los índices para que $y_t = D_t$ esté alineado con el conjunto de fechas correspondiente.
- Ajustamos el modelo
$$D_t = f_\theta(t) + \varepsilon_t.$$

Tras ajustar, tenemos un estimador suave de las diferencias diarias:

$$\hat{D}_t = f_{\hat{\theta}}(t).$$

5. "Destransformación": reconstruir un ***trend*** en el espacio de precios

El punto clave: **sólo destransformamos la salida del modelo $\hat{y}_t$, no la serie observada y_t.**

5.1. Transformaciones directas (niveles, logaritmos, etc.)

Cuando g es biyectiva punto a punto $P_t \mapsto y_t$, por ejemplo:

- `close`: $y_t = P_t$, entonces $h(y_t) = y_t$.
- `log_close`: $y_t = \log P_t$, entonces $h(y_t) = e^{y_t}$.

Definimos una inversa h tal que

$$h(g(P_t)) = P_t.$$

En este caso, el *trend* de precios que dibujamos es

$$\hat{P}_t = h(\hat{y}_t),$$

y la banda de confianza en el espacio de precios es

$$\hat{P}_t^{\text{low}} = h\big(\hat{y}_t + q_{\text{low}}\big), \qquad \hat{P}_t^{\text{high}} = h\big(\hat{y}_t + q_{\text{high}}\big),$$

donde $q_{\text{low}}, q_{\text{high}}$ son cuantil(es) empíricos de los residuos en el *espacio transformado*.

5.2. Transformaciones acumulativas (diferencias, rendimientos)

Para `diff_close`, el transform es

$$y_t = D_t = P_t - P_{t-1}.$$

Ya no existe una inversa punto a punto h que dependa sólo de y_t; la información de P_t está codificada *acumulativamente* en todos los incrementos previos. La relación verdadera es

$$P_t = P_0 + \sum_{s=1}^{t} D_s.$$

En el modelo, sustituimos D_s por el valor ajustado $\hat{D}_s$:

$$\hat{P}_t = P_0 + \sum_{s=1}^{t} \hat{D}_s = P_0 + \sum_{s=1}^{t} f_{\hat{\theta}}(s).$$

En código se implementa como recursión:

$$\begin{aligned} \hat{P}_0 &= P_0, \\ \hat{P}_t &= \hat{P}_{t-1} + \hat{D}_t, \qquad t = 1, \dots, n-1. \end{aligned}$$

Esto produce un **trend suave** en el espacio de precios:

$$\hat{P}_t = \text{“precio suavizado”}$$

porque $\hat{D}_t$ es una función suave de t (polinomio regularizado).

Análogamente, si usamos una banda empírica en el espacio de diferencias:

$$\hat{D}_t^{\text{low}} = \hat{D}_t + q_{\text{low}}, \qquad \hat{D}_t^{\text{high}} = \hat{D}_t + q_{\text{high}},$$

reconstruimos las bandas de precio por

$$\begin{aligned} \hat{P}_0^{\text{low}} &= P_0, \qquad \hat{P}_t^{\text{low}} = \hat{P}_{t-1}^{\text{low}} + \hat{D}_t^{\text{low}}, \\ \hat{P}_0^{\text{high}} &= P_0, \qquad \hat{P}_t^{\text{high}} = \hat{P}_{t-1}^{\text{high}} + \hat{D}_t^{\text{high}}. \end{aligned}$$

Conclusión importante. En el caso `diff_close`:

- Nunca reconstruimos P_t usando los verdaderos incrementos D_t en la gráfica de *trend*.
- Siempre usamos los *incrementos ajustados* $\hat{D}_t$ para obtener una curva suave $\hat{P}_t$.
- La serie observada P_t se dibuja tal cual, con sus brincos; la curva $\hat{P}_t$ es un suavizado.

Por eso, visualmente, la línea negra (trend) es continua y suave, mientras que la serie de puntos/segmentos de la serie cruda puede ser más “salto a salto”.

6. Qué se destransforma y qué no

Podemos resumir así:

1. **Espacio transformado.**
   - Datos observados: $y_t = g(P_\bullet)$.
   - Modelo: $\hat{y}_t = f_{\hat{\theta}}(t)$.
   - Residuos: $r_t = y_t - \hat{y}_t$.
2. **Espacio de precios (destransformado).**
   - Serie observada: P_t se dibuja tal cual (`df[çlose"]`).
   - **Sólo destransformamos:**
     - La curva de trend $\hat{y}_t \mapsto \hat{P}_t$.
     - Las bandas de confianza $\hat{y}_t + q \mapsto \hat{P}_t^{(\cdot)}$.

En particular, nunca hacemos

$$y_t \xrightarrow{h} P_t^\star$$

para luego sustituir P_t por $P_t^\star$; es decir, no "destransformamos" la serie observada, sólo la salida del modelo.

7. Ejemplo compacto: `diff_close`

Para fijar ideas, el pipeline de `diff_close` es:

$$\begin{aligned}
D_t &= P_t - P_{t-1}, \\
D_t &= f_\theta(t) + \varepsilon_t, \\
\hat{D}_t &= f_{\hat{\theta}}(t), \\
\hat{P}_0 &= P_0, \\
\hat{P}_t &= \hat{P}_{t-1} + \hat{D}_t.
\end{aligned}$$

Visualmente:

- En la gráfica de **transformación** (`diff_close`):

$$\boxed{D_t \text{ observado (salta)} \quad vs \quad \hat{D}_t \text{ suave.}}$$

- En la gráfica de **precio**:

$$\boxed{P_t \text{ observado (salta)} \quad vs \quad \hat{P}_t = P_0 + \sum_{s=1}^{t} \hat{D}_s \text{ suave.}}$$

8. Resumen simbólico

La idea central que explica tu observación sobre la línea continua es:

$$\boxed{\begin{array}{l} (1) \text{ Transformamos la serie: } y_t = g(P_\bullet). \\ (2) \text{ Ajustamos } y_t \approx f_\theta(t) \;\Rightarrow\; \hat{y}_t. \\ (3) \text{ Construimos un } \textit{trend} \text{ en precios: } \hat{P}_t = H(\hat{y}_{0:t}), \\ \quad \text{donde } H \text{ integra o invierte } g \text{ usando sólo la salida del modelo.} \end{array}}$$

En el caso de diferencias, H es la suma acumulada de incrementos ajustados; como éstos son suaves, la curva $\hat{P}_t$ es también suave y no reproduce los saltos originales de P_t, sino su tendencia estimada.