Rough Paths:

- Motivación original $\leftarrow$ estudiar "ecuaciones diferenciales" que incluyen perturbaciones aleatorias.

- Motivación reciente $\leftarrow$ Realizar aprendizaje de máquina interpretando datos como trayectorias de funciones continuas.

Motivación clásica:

$Y_t = f(t)$ $\leftarrow$ Función del tiempo.

$$\frac{dY_t}{dt} = V(Y_t) \quad \text{con} \quad Y_0 = y_0 \qquad V: \mathbb{R} \to \mathbb{R}.$$

$$\downarrow$$

$$dY_t = V(Y_t)\,dt \quad \text{con} \quad Y_0 = y_0$$

$$\Updownarrow$$

$$Y_t - Y_s = \int_s^t V(Y_r)\,dr$$

Siguiente paso de generalidad:

$$dY_t = V(Y_t)\,dX_t$$

$X_t \leftarrow$ otra función del tiempo que satisface ciertas condiciones.

$$\Updownarrow$$

$$Y_t - Y_s = \int_s^t V(Y_r)\,dX_r$$

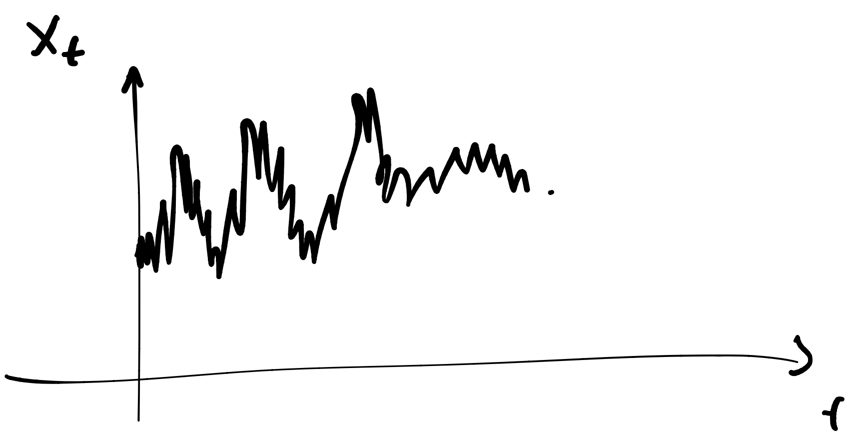
$$\underbrace{dY_t}_{\text{Output}} = V(Y_t)\,\underbrace{dX_t}_{\text{input.}}$$

¿Aplicación? X_t = precio en pesos del dolar.

Y_t = Cartera de inversión de divisas.

Lado negativo:

Si vemos el mercado en tiempo real:



$\leftarrow$ Podría ser un problema.

Análisis detallado del problema:

$$dY_r = V(Y_r)\, dX_r \iff Y_{r+\varepsilon} - Y_r \approx V(Y_r)\,(X_{r+\varepsilon} - X_r).$$

Samplear X en $[0,1]$ en tiempos

$t_1 = \frac{1}{n}$, $t_2 = \frac{2}{n}$, . . . , $t_n = 1 = \frac{n}{n}$.

Si $X_i = X_{t_i}$ $Y_i = Y_{t_i}$

$$Y_{i+1} = Y_i + V(Y_i)\,\Delta_i X$$

$\Delta X_t \approx \Delta t$ + error

$$Y_{i+1} = Y_i + V(Y_i)\left(\frac{1}{n} + \frac{1}{\sqrt{n}}\xi_i\right).$$

modelo

$\{\xi_i\} \leftarrow$ v.a.i.i.d. $\xi_i \sim \mathcal{N}(0,1)$.

Cuando tomamos límite obtenemos

$$dY_t = V(Y_t)\,(dt + dB_t).$$

$$Y_t - Y_s = \int_s^t V(Y_r)\,dr + \underbrace{\int_s^t V(Y_r)\,dB_r}_{\text{es "sutil" de definir.}}$$

Esto motiva a formular ecuaciones del tipo

$$dY_t = V(Y_t)\,dX_t.$$

$X_t \leftarrow$ Potencialmente tiene trayectorias muy erráticas.

¿Cómo depende el output en función del input?

$\updownarrow$

¿continuamente?

$$dY_t = V(Y_t)\,dX_t.$$

Proposición (Lyons et. al).
No existe un espacio de Banach $\mathcal{B}$ t.q. las trayectorias de un mov. Browniano pertenezcan a $\mathcal{B}$ casi seguramente y t.q. el mapeo

$$(f,g) \longmapsto \int_0^{(\cdot)} f(t)\,\dot{g}(t)\,dt,$$

definido primero sobre f, g diferenciables $f,g:\mathbb{R}\to\mathbb{R}$ se extiende a $\mathcal{B}\times\mathcal{B}$ de manera continua.

Como esquivar esquivar esta restricción:

$$Y_t - Y_s = \int_s^t V(Y_r)\, dX_r \quad *$$

Framework. general:

$$Y_t = (Y_t^1, \dots, Y_t^e) \qquad e \in \mathbb{N}.$$

$$X_s = (Y_s^1, \dots, Y_s^d). \qquad d \in \mathbb{N}.$$

$$V \in \mathbb{R}^{e \times d} \qquad V = (V_{i,k} \,;\, 1 \le i \le e, \quad 1 \le k \le d).$$

$$Y_t^i - Y_s^i = \sum_{k=1}^{d} \int_s^t V_{i,k}(Y_r)\, dX_r^k \quad \leftarrow \text{nuestra ecuación}$$

se suponer que tengo acceso a Y_s, X_s.

$$Y_t^i - Y_s^i = \sum_{k=1}^{d} \int_s^t \left((V_{i,k}(Y_s)) + (V_{i,k}(Y_r) - V_{i,k}(Y_s)) \right) dX_r^k$$

$$= \sum_{k=1}^{d} V_{i,k}(Y_s) \underbrace{\int_s^t dX_r^k}_{\text{Importante}}$$

$$+ \sum_{k=1}^{d} \int_s^t \sum_{i_2=1}^{e} \int_s^r \frac{\partial V_{i,k}}{\partial y_{i_2}} (Y_{r_2})\, dY_{r_2}^{i_2}\, dX_r^k$$

$$\int_s^t \int_s^r \frac{\partial V_{i,r}}{\partial y_{i_2}}(Y_{r_2})\, dY_{r_2}^{i_2}\, dX_r^{K}$$

$$= \int_s^t \int_s^r \frac{\partial V_{i,r}}{\partial y_{i_2}}(Y_{r_2}) \sum_{K_2=1}^{d} \frac{\partial V_{i_2,K_2}}{\partial y_{i_2}}(Y_{r_2})\, dX_{r_2}^{K_2}\, dX_r^{K}$$

$$= \int_s^t \int_s^r \frac{\partial V_{i,r}}{\partial y_{i_2}}(Y_s) \sum_{K_2=1}^{d} \frac{\partial V_{i_2,K_2}}{\partial y_{i_2}}(Y_s)\, dX_{r_2}^{K_2}\, dX_r^{K}$$

+ Terminos similares.

$$= \sum \text{Coeficientes}(s) \int_s^t \int_s^{r_1} \cdots \int_s^{r_l} dX_{r_{l+1}}^{K_{l+1}} \cdots dX_{r_1}^{K_1}$$

Definimos

$$S_{s,t}^{K_1,\dots,K_l}[x] := \int_s^t \cdots \int_s^{r_{l-1}} dX_{r_l}^{K_l} \cdots dX_{r_1}^{K_1}$$

Ahora, en lugar de tomar el mapeo:

$$\{X_t\}_{t\geq 0} \xrightarrow[\text{continuo}]{\text{No es}} \{Y_t\}_{t\geq 0}$$

$$\{X_t\}_{t\geq 0} \xrightarrow[\text{agregar info}]{\text{medible}} \left(\{X_t\}_{t\geq 0},\ \text{todas las } S_{s,t}^{k_1,\dots,k_\ell}[X]\right)$$

hasta $\ell \leq q$.

$$\Big\downarrow \text{ si es continuo.}$$

si q es suficientemente grande (depende de la Hölder continuidad de X

si X es α-Hölder continuo.

$$q = \lfloor \tfrac{1}{\alpha} \rfloor$$

$$\{Y_t\}_{t\geq 0}.$$

Aplicaciones en machine learning:

Objeto de interés:

Signaturas $\to \left\{ S_{s,t}^{k_1,\dots,k_q}[X] \right\}$. suponer $s=0$, $t=1$

Remark: Signaturas son "características" especiales de $\{X_t\}_{t\geq 0}$.

Problema práctico:

- $Y_t^1 = \alpha + \beta_1^1 Y_{t-1}^1 + \beta_1^2 Y_{t-2}^2 + \varepsilon_t^1 \leftarrow$ clasif. 1

- $Y_t = \alpha + \beta_2^1 Y_{t-1}^1 + \beta_1^2 Y_{t-2}^2 + \varepsilon_t^2 \leftarrow$ clasif. 2.

Problemas de clasificación que se pueden resolver con técnicas de rough path.

Analogía con regresión:

$Y \leftarrow$ respuesta (score crediticio).

$$Y = \beta_0 + \sum_{j=1}^{d} \beta_j X_j \qquad X_i \leftarrow \text{características}, \quad X_i \in \mathbb{R}.$$

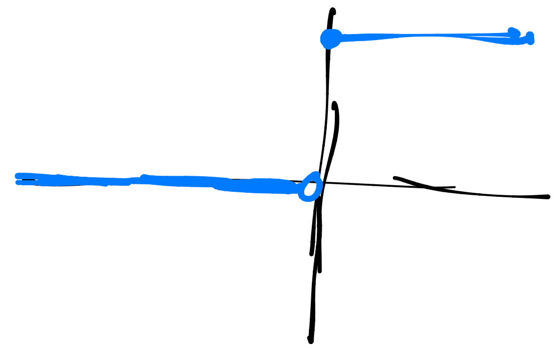
$$\left\{ \begin{array}{l} Y^1, \dots, Y^n \\ (X_1^1, \dots, X_d^1), \dots, (X_1^n, \dots, X_d^n) \end{array} \right. \leftarrow \text{observaciones}$$

$\beta_0, \dots, \beta_d \leftarrow$ tienen una fórmula para calcularse.

Pasamos ahora de regresión a clasificación

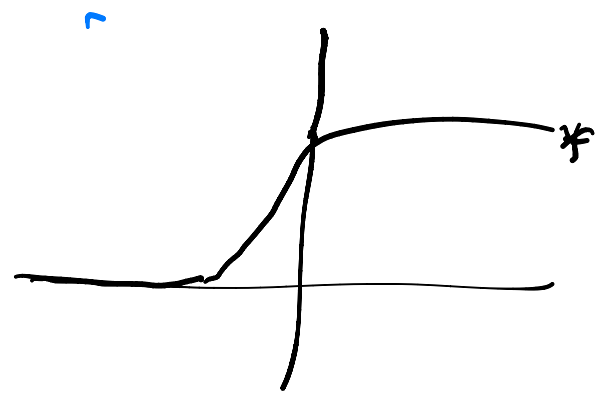
Me gustaría:

$$Y = \phi\left(\beta_0 + \sum_{j=1}^{d} \beta_j X_j\right)$$



$\phi(x) = \mathbb{1}_{\{x \geq 0\}}$

$$Y = \varphi\left(\beta_0 + \sum_{j=1}^{d} \beta_j X_j\right)$$



$\phi(x) =$ sigmoide como en

Nos da un aprox. de la clasificación.

Nota:

Si queremos reducir

$$\sum_{\ell=1}^{n} \mathcal{E}\left(Y^\ell - \varphi\left(\beta_0 + \sum_{j=1}^{d} \beta_j X_j^\ell\right)\right) + \sum_{j=1}^{d} w_j \tilde{\mathcal{E}}(\beta_j)$$

$\mathcal{E} \leftarrow$ cuantificación del error $\qquad \tilde{\mathcal{E}} \leftarrow$ penaliz.

De vuelta al problema de clasificación de series de tiempo:

- $X_t^1 = \alpha + \beta_1^1 X_{t-1}^1 + \beta_1^2 X_{t-2}^2 + \varepsilon_t^1 \leftarrow$ clasif. 1 #

- $X_t = \alpha + \beta_2^1 X_{t-1}^1 + \beta_1^2 X_{t-2}^2 + \varepsilon_t^2 \leftarrow$ clasif. 2. ##

Queremos:

- $Y \leftarrow$ respuesta: eres # o ## ✓

Valores de $Y \leftarrow \{0, 1\}$. ✓

Ahora vamos a observar:

Voy a ver algo como:

Input $\leftarrow$ la trayectoria de la serie

Problema: como aplico lo que tengo en slides anteriore ?

$\downarrow$ traducción:

Obtener "características" extraídas de la trayectoria de lo serie de tiempo?

$$X \xmapsto{\tilde{S}} \underbrace{(\tilde{S}_1(X), \dots, \tilde{S}_{40}(X))}_{\text{características.}}$$

Ahora: puedo proponer números $\beta_0, \dots, \beta_4$ t.q.

$$Y = \varphi\left(\beta_0 + \sum_{i=1}^{40} \beta_i \tilde{S}_i(X)\right) + \text{error de medición.}$$

Ésta es "toda" la metodología, siempre y cuando $\tilde{S}_i$ ya estén elegidas,

Conexión con Rough paths:

Comentario:

Hay info. que me gustaría que no influyera en la variable de respuesta.

- Reconocimiento de acciones ← Marcos de referencia no importan

- Reconocimiento de caractéres chinos. ← invarianza bajo "dibujar más rápido".

Caracteristicas: $X \in \mathbb{R}^d$

$$(\tilde{S}_1(X), \dots, \tilde{S}_{40}(X))$$

$$= (S_{i_1,\dots,i_q}(X) ; \quad 1 \le q \le Q \quad i_j \in \{1,\dots,d\})$$

Justificación de que $S_{\underset{\sim}{i}}$ son "buenas candidatas" ← depende.

- Ventaja: 1 $S_{i_1,\dots,i_q}(x)$ son invariantes bajo reparametrizaciones.

- Ventaja 2:

$$(S_{i_1,\dots,i_q} \ ; \ q \in \mathbb{N}\setminus\{0\},\ i_j \in \{1,\dots,d\}).$$

$$\downarrow$$

$$\tilde{T} := \bigoplus_{q=1}^{\infty} (\mathbb{R}^d)^{\otimes q} \longleftarrow = \left\{ \sum_{q=1}^{\infty} \sum_{i_1,\dots,i_q=1}^{d} a_{i_1,\dots,i_q}\, e_{i_1} \otimes \cdots \otimes e_{i_q} \right\}$$

Para encajar a las $S_{i_1,\dots,i_q}$ en $\tilde{T}$

escogemos $a_{i_1,\dots,i_q} = S_{i_1,\dots,i_q}$.

Ventaja:

Si consideramos

$$\int_s^{t}\int_s^{t_1}\int_s^{t_2}\cdots\int_s^{t_q} 1\, dX_{t_q} \otimes dX_{t_{q-1}} \otimes \cdots \otimes dX_1 \in (\mathbb{R}^d)^{\otimes q}$$

$$\downarrow$$

$$:= \sum_{i_1,\dots,i_q} \left(\int_s^{t}\int_s^{t_1}\int_s^{t_2}\cdots\int_s^{t_q} 1\, dX_{t_q}^{i_q}\, dX_{t_{q-1}}^{i_{q-1}} \cdots dX_1^{i_{q-1}} \right) e_{i_1} \otimes \cdots \otimes e_{i_q}$$

Nota:

El "espacio de Fock" $\tilde{F} = \bigoplus_{q=1}^{\infty} (\mathbb{R}^d)^{\otimes q}$

tiene un producto que le llamamos "$\otimes$"

$$\underset{\substack{\Uparrow \\ \tilde{F}}}{(e_{i_1} \otimes e_{i_2})} \text{"}\otimes\text{"} \underset{\substack{\Uparrow \\ \tilde{F}}}{(e_{i_3})} = e_{i_1} \otimes e_{i_2} \otimes e_{i_3}$$

Ventaja 2 :)

Por inv. bajo reparametrización:

puedo pensar que X_t solo se considera

para $t \in [0,1]$.

$Y_t \leftarrow$ otra "trayectoria" $t \in [1,2]$ t.q. $Y_1 = X_1$

Tenemos un producto en las trayectorias.

$$(X \underset{\substack{\uparrow \\ \text{producto en las trayectorias.}}}{\odot} Y)_t = \begin{cases} X_t & \text{si } t \in 1 \\ Y_t & \text{si } t \in [1,2] \end{cases} \qquad t \in [0,2]$$

$$S(X \odot Y) = S(X) \otimes S(Y) \leftarrow$$ Signaturas pueden calcularse en sub-intervalos y luego "pegarse"

Ventaja 2.2)

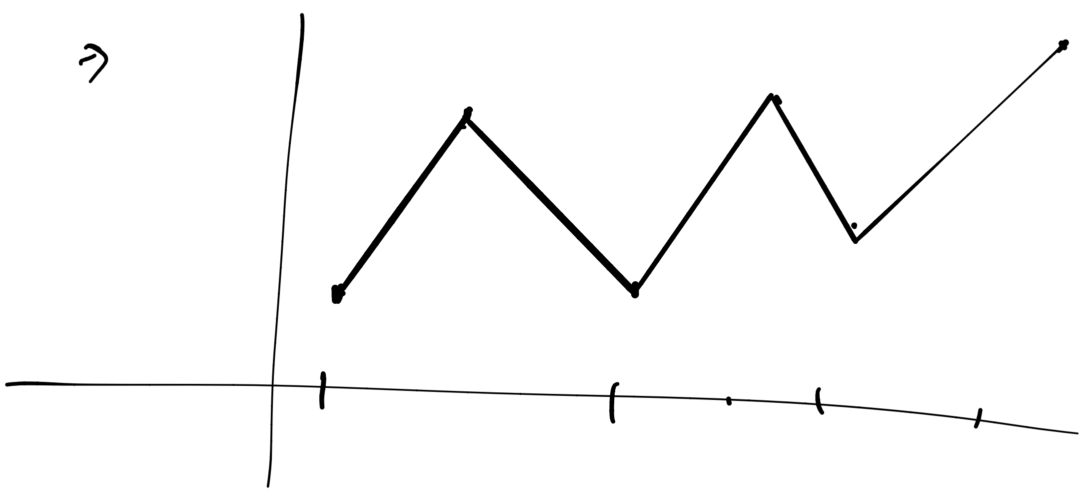
$X_t = t\,v \qquad v \in \mathbb{R}^d$

Entonces

$$1 + \int_0^1 dX_v + \int_0^1 \int_0^{t_1} dX_{t_2} \otimes dX_{t_1} + \dots$$

$$= 1 + \sum_{k=1}^{M} \frac{v^{\otimes k}}{k!} = \exp\{v\}.$$

$\Rightarrow$



X_t ← diferenciable.

$$\int_0^1 X_t \, dX_t$$

$$S_{i_1,\dots,i_{q_1}} S_{j_1,\dots,i_{q_2}} = \sum_{(w_1,\dots,w_{q_1+q_2}) \in \mathcal{I}} \boxed{} \, S_{w_1,\dots,w_{q_1+q_2}}$$

C_q = q'ésimo coef de $\log(\mathbb{E}[\exp\{\lambda X\}])$.

$d=1$ en $[0,1]$

candidato de

$S_{t_1,\dots,t_q} =$ Caos de orden q. B.

$$e^{-\frac{\lambda}{2} B_0 - \frac{\lambda^2}{2} \square}$$