Análisis Matemático para Inteligencia Artificial

Martín Errázquin (merrazquin@fi.uba.ar)

Especialización en Inteligencia Artificial

Exponential Moving Averages (EMA)

EWMA/EMA/Exponential Smoothing

Idea: quiero ir promediando (o algo parecido) pero darle más peso a lo reciente para que lo muy viejo no condicione tanto.

Dada una sucesión de valores a medir $x_1, \ldots, x_t$ el λ-*Exponentially Weighted Moving Average* es:

suc. de "mediciones"

suc. de "estimaciones"

$$a_t = \lambda \cdot a_{t-1} + (1 - \lambda) \cdot x_t$$

$\lambda = 0.5$

Podemos ver que, aproximadamente, a_n toma la forma de $\sum_{i=1}^{n} \lambda^{n-i} x_i = x_n + \lambda x_{n-1} + \lambda^2 x_{n-2} + \lambda^3 x_{n-3} + \ldots$. Usando $\lambda \in (0, 1)$ es un promedio ponderado que favorece valores recientes.

1   0.5   0.25   0.125

1   0.9   0.81   0.729

$\lambda = 0.9$

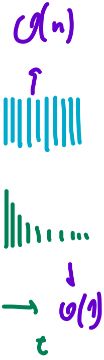
Y algo muy importante: es $\mathcal{O}(1)$ en memoria y $\mathcal{O}(1)$ para actualizar!

Nota: El valor inicial a_0 es un hiperparámetro, aunque es bastante común usar 0, la media global de x ($\overline{x}$) o directamente el primer valor x_1.

$\lambda = 0.8$ $x_1 = 2 \Rightarrow a_1 = 0.8 \cdot 0 + 0.2 \cdot 2 = \underline{0.4}$ $x_3 = 2 \Rightarrow a_3 = 0.8 \cdot 0.52 + 0.2 \cdot 2 = 0{,}816$

$a_0 = 0$ $x_2 = 1 \Rightarrow a_2 = 0.8 \cdot \underline{0.4} + 0.2 \cdot 1 = 0.52$

EWMA vs Windowed-Average (SMA)



A. Kofman http://spotware.ctrader.com/c/XKB5n, https://commons.wikimedia.org/w/index.php?curid=27263134

Rescaling

Formato de promedio común es como ponderar todo con $\frac{1}{n}$:

prom. de $x, x, x, \ldots, x$

$\frac{\sum_{i=1}^n x_i}{n} = \frac{1}{n}\sum_{i=1}^n x_i = \sum_{i=1}^n \frac{1}{n} x_i$. Importante: $\sum_{i=1}^n \frac{1}{n} = 1$.

$\Rightarrow x$, no $\frac{x}{10}$

¿Qué pasa con el EWMA? Para una sucesión que inicia en $t = 0$ con $a_{-1} = 0$, tenemos $a_n = \sum_{i=0}^n \lambda^{n-i}(1-\lambda)x_i$. Veamos que:

$$\lambda^0 + \lambda^1 + \cdots + \lambda^n = \sum_{i=0}^n \lambda^i = \frac{1-\lambda^n}{1-\lambda}$$

Entonces,

$(1-\lambda^n)x$ $\in (0,1)$

$$\sum_{i=0}^n \lambda^{n-i}(1-\lambda) = \sum_{i=0}^n \lambda^i(1-\lambda) = \frac{1-\lambda^n}{1-\lambda}(1-\lambda) = 1-\lambda^n$$

La suma de los pesos no da 1! Si $n >> 0$ entonces $1-\lambda^n \approx 1$, pero en n chicos se nota mucho. Por ej. $\lambda = 0.99, n = 10 \Rightarrow 1-\lambda^n \approx 0.0956$.
La solución? Reescalar dividiendo por $1-\lambda^n$.

Rescaling: ejemplo simple

$a_0 = b_0 = 0$

Sea una sucesión de valores $x_n = 1, 1.5, 2, 1.6, 4, 2, 1.3, \ldots$ y $\lambda = 0.99$
¿Cómo cambian los EMA si se usa o no rescaling? ¿Y si $\lambda = 0.9$?

med. EMA(λ=0.99) EMA (λ=0.9)

x_n	a_n	$1-0.99^n$	a'_n	b_n	$1-0.9^n$	b'_n
1	0.01	0.01	1.0	0.1	0.1	1.0
1.5	0.0249	0.02	1.25	0.24	0.19	1.26
2	0.0447	0.03	1.5	0.416	0.271	1.54
1.6	0.0602	0.039	1.53	0.534	0.3439	1.55
4	0.0996	0.049	2.03	0.881	0.4095	2.15
2	0.119	0.059	2.03	0.993	0.4686	2.12
1.3	0.13	0.068	1.92	1.02	0.5217	1.96