Modelagem de Séries Temporais

curso-r

Modelos AR

- No nosso exemplo de suavização exponencial temos a fórmula:

$$\hat{y}(t) = \alpha y(t) + (1-\alpha)\hat{y}(t-1)$$

Como a mesma fórmula funciona para $t-1$, também temos que:

$$\hat{y}(t-1) = \alpha y(t-1) + (1-\alpha)\hat{y}(t-2)$$

Substituindo fica:

$$\hat{y}(t) = \alpha y(t) + (1-\alpha)\alpha y(t-1) + (1-\alpha)^2\hat{y}(t-2).$$

Logo, estamos construindo uma previsão $\hat{y}$ que é a média das anteriores, mas uma média que dá muito mais peso para a primeira do que as anteriores.

Modelos AR

- De fato, modelos com essa cara são importantes e recebem até um nome específico: modelos auto-regressivos. Nesses modelos, a equação da nossa previsão é parecida com:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2}$$

Claro que podemos pegar muitos dados para trás:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots \phi_p y_{t-p}$$

Normalmente admitimos que essa equação tem algum tipo de erro, então, parecido com o caso de uma regressão, dizemos que o modelo tem um erro ϵ_t ruído branco:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots \phi_p y_{t-p} + \epsilon_t$$

Dizemos que esse modelo é Auto-Regressivo de ordem p ou $AR(p)$

Exemplos de modelos AR

Modelos AR normalmente desenham retinhas em que pontos próximos são parecidos:

escolhendo $p = 1$ e $\phi_1 = 0.9$

Exemplos de modelos AR

Note como esse desenho é diferente do ruído branco:

Inclusive, o ruído branco aparece quando escolhemos $p = 1$ e $\phi_1 = 0$

Exemplos de modelos AR

Exemplo muito especial:

O gráfico abaixo, um AR com $p = 1$ e $\phi_1 = 1$, te lembra alguma coisa?

Porque o modelo auto regressivo é importante?

- Alguns fenômenos são quase que perfeitamente modelos AR, como o preço de uma ação, por exemplo.
- Gráficos de ACF e PACF, que as vezes são meio difícies de interpretar, tem padrões específicos no modelo.
- O AR(1), por exemplo, tem uma ACF bem especial
- Modelos auto-regressivos costumam ter ACF que decaem rápido, mas ficam um bom tempo longe do zero. No caso do AR(1), a ACF tem até fórmula:

$$ACF(t) = \phi_1^t$$

ACF do AR(1)

Vamos pensar no porque isso é desse jeito com $\phi_1 = .8$

ACF de um AR com coeficiente negativo

Quando ϕ é negativo fica mais estranho:

$\phi_1 = -.8$

PACF

Justamente pensando no impacto "persistente" do passado na série como um todo, foi invetado o conceito de **F**unção **A**uto **C**orrelação **P**arcial. Ela é que nem a auto correlação, mas ao invés de fazer a correlação diretamente, ela "limpa" os efeitos persistentes fazendo regressões.

Por exemplo, vamos pensar no modelo AR(1):

$$y_t = y_{t-1} + \epsilon_t$$

$$y_{t-1} = y_{t-2} + \epsilon_{t-1} \implies y_{t-2} = y_{t-1} - \epsilon_{t-1}$$

Como tanto o y_t quanto o y_{t-2} podem ser escritos em função de y_{t-1}, é natural que eles sejam correlacionados: eles são montados a partir do mesmo número. O que a PACF faz é calcular a correlação entre ϵ_{t-1} e ϵ_t, esses sim são não rorrelacionados.

PACF na prática

A PACF na prática nos ajuda a identificar o grau adequado de autocorrelação de uma série. Note como é a forma da PACF de um AR(1) com $\phi_1 = .8$

PACF na prática

PACF de um AR(2) e $\phi_1 = \mathbf{.7}$ e $\phi_2 = \mathbf{.3}$

PACF na prática

PACF na prática

ACF com bastante autocorrelação:

PACF na prática

PACF consegue identificar que muita dessa correlação é reduntante, temos uma autocorrelação no primeiro grau muito alta e talvez uma autocorrelação de ordem maior.

Como modelos alterar um modelo AR

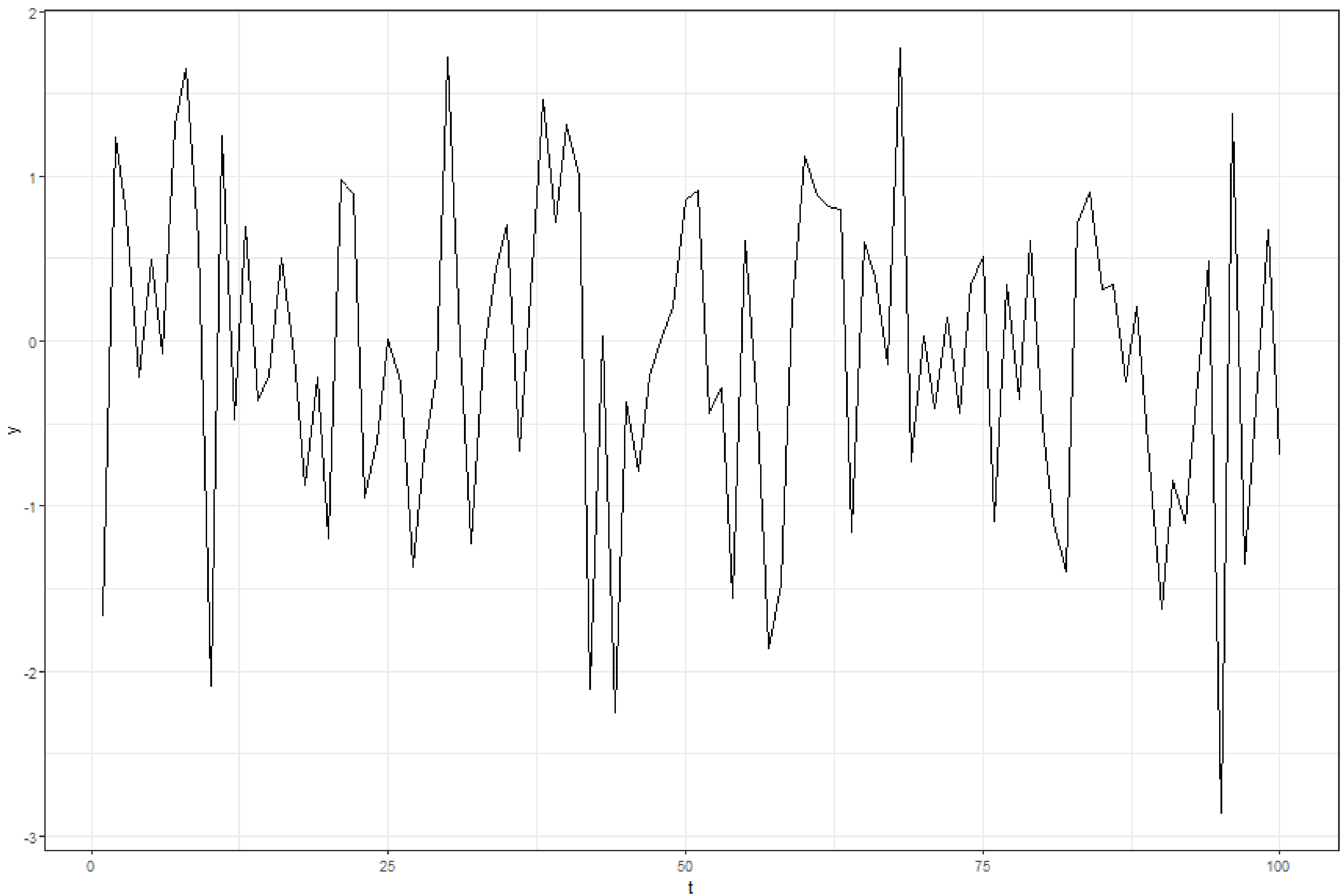
Pensar em um AR ajuda a interpretar auto-correlações, mas as PACFs típicas de um modelo AR são muito rígidas. A PACF da série da ANAC não concorda com um AR simples:

Processo MA

Uma média móvel usa uma ideia parecida com a do AR para montar um modelo que se parece com o ruído branco, mas é varia menos em janelas próximas

$$y_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \ldots + \theta_q \epsilon_{t-q}$$

Ruído branco (de novo)



Média movel

Vamos escolher $q = 1$ e $\theta_1 = 0.9$

ACF do modelo MA(q)

$q = 1$ e $\theta_q = 0.9$

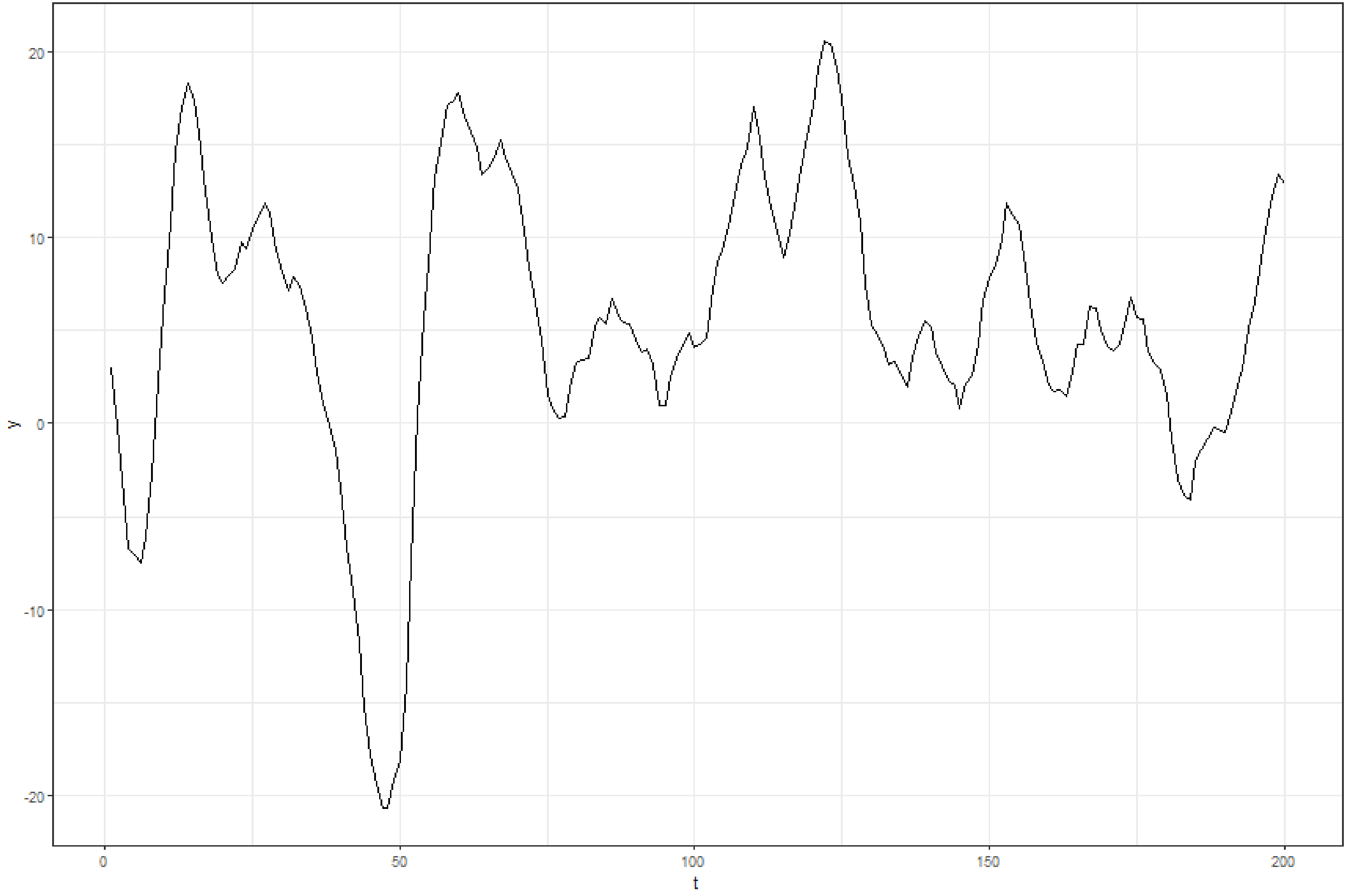
ACF do modelo MA(q)

$q = 2$ e $\theta_1 = 0.9, \theta_2 = 0.9$

Processo ARMA

Um processo ARMA é um processo que mistura a parte autoregressiva com o resíduo média móvel, que pode ter correlações em um janela específica. Isso garante que o "desenho" de um processo arma é menos dentado do que um modelo auto regressivo simples. Considere essa simulação de um ARMA(1,1):

ARMA(1, 6)



Intuição geral

Um modelo ARMA é um modelo:

- De caráter **Auto Regressivo**: y_t, y_{y-1}, y_{t-2} etc podem ter relação direta e persistente entre eles.
- Com resíduos em **Média Movel**: y_{t-1} vai a y_t de maneira suave, similar ao jeito que y_{t-2} foi a y_{t-1}

Aparentemente olhando todos os modelos $ARMA(p, q)$ temos uma grande ferramenta para ajustarmos praticamente qualquer série temporal, mas faltam dois elementos **tendência** e **sazonalidade**.

Diferenças

Um jeito rudimentar de "limpar" uma série temporal de tendência e possivelmente de sazonalidade é através do método das diferenças. Isso é, criar uma nova série temporal d da seguinte forma:

$$d_t = y_t - y_{t-1}$$

Voltando à série da ANAC

Essa série tem tendência e possivelmente também tem sazonalidade:

Diferenças da série da ANAC

Aqui a tendência sumiu:

ACF da série de Diferenças da ANAC

PACF da série de Diferenças da ANAC

Diferença da diferença

Aqui a série talvez já possa ser modelada adequadamente por um ARMA...

ACF da série de Diferenças da ANAC

Modelo ARIMA

Assim como sugerimos no último exemplo, um modelo ARIMA nada mais é que um modelo ARMA (auto regressivo e com resíduo em média móvel) ajustado à uma (ou mais) diferenças de uma série temporal. Justamente por essa característica, fora os pesos, um modelo ARIMA costima vir acompanhando de três letras:

- p que vem do $AR(p)$,
- d que vem da ordem da diferença (se é diferença simples, ou diferença da diferença, ou diferença da diferença da diferença etc)
- q que vem do $MA(q)$

Logo, um $ARIMA(p, 1, q)$ seria:

$$y_t - y_{t-1} = d_t \sim ARMA(p, q)$$

Já um $ARIMA(p, 2, q)$ seria:

$$d_t - d_{t-1} = y_t - y_{t-1} - y_{t-1} + y_{t-2} \sim ARMA(p, q)$$