Tópicos Avançados em IA

Prof. James Moraes de Almeida

ilum escola de ciência

CNPEM
Centro Nacional de Pesquisa em Energia e Materiais



Cronograma

- Séries temporais:
  - 01/10/2026: O que são e decomposição.
  - 08/10/2026: Previsão.
  - 22/10/2026: Métodos de aprendizado de máquina.
- Aprendizado ativo e otimização bayesiana:
  - 29/10/2026: Aprendizado ativo.
  - 05/11/2026: Processos gaussianos e otimização bayesiana.
- 19/11/2026: Apresentações em duplas.

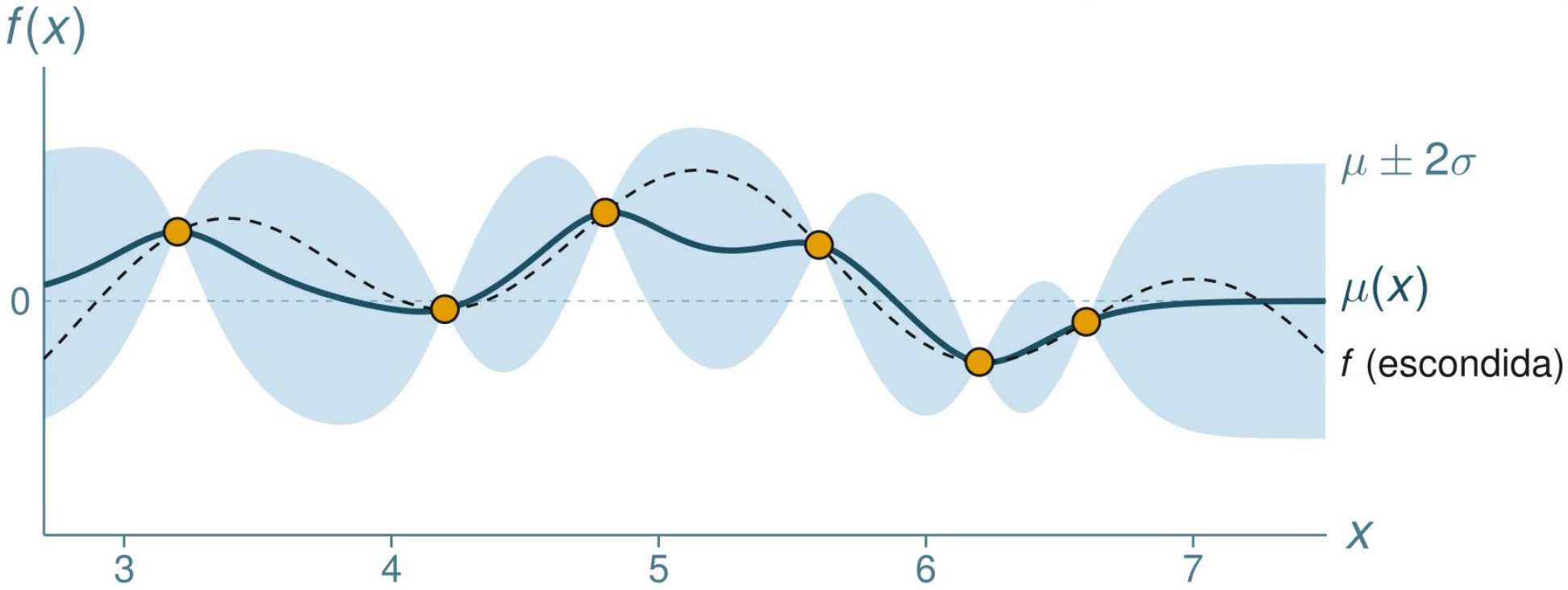
Otimização Bayesiana

- Da Aula 04 para hoje
  - Aula 04: qual ponto rotular para o modelo ficar bom?
  - Hoje: onde está o máximo, se cada avaliação é cara?
    - O aço mais resistente (o enigma da Aula 04), o melhor hiperparâmetro de um modelo.
- Precisamos de um modelo que saiba o quanto não sabe.

Otimização Bayesiana

Sorteando funções

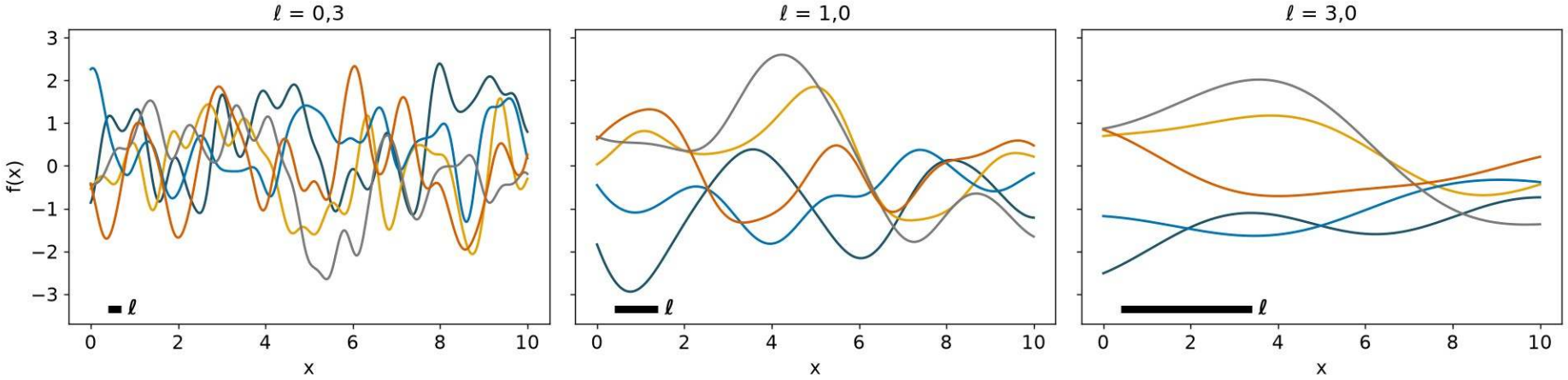
$$f(x) = -\left(\operatorname{sen} x + \operatorname{sen}(10x/3)\right)$$



Com palpite e dúvida, decidimos onde medir.

Otimização Bayesiana

- Processo gaussiano (GP): cada sorteio é uma função

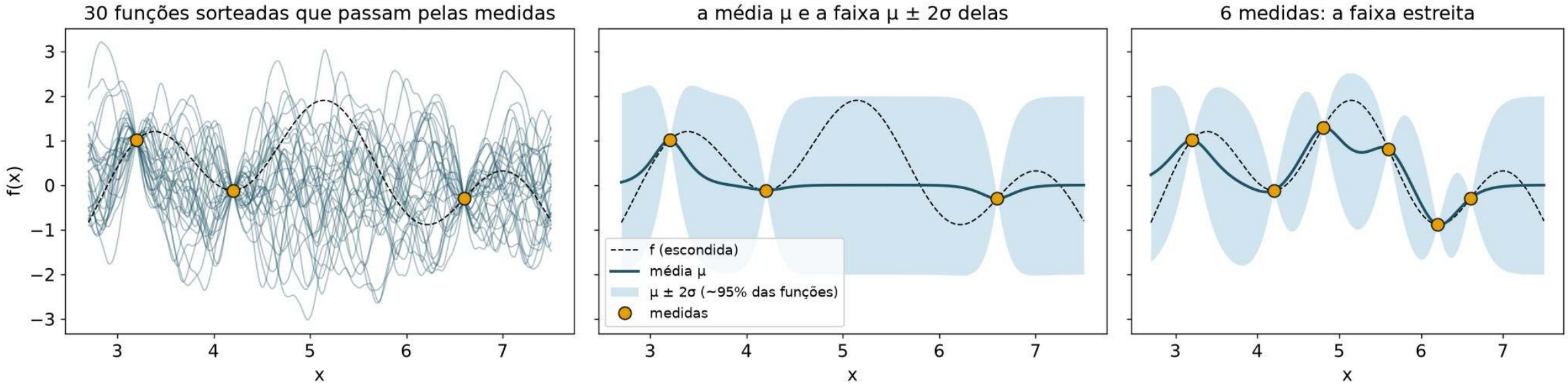


$$k(x, x') = \sigma^2 \exp\left(-\frac{(x - x')^2}{2\ell^2}\right)$$

- x': onde já medimos · x: um ponto ainda sem medida
- σ^2: o quanto f varia (as curvas ficam entre ±2σ)
- ℓ: comprimento de correlação — a uma distância ℓ, a correlação é 0,61.
- Com $x' = x$, $k = \sigma^2$; com x' a 3ℓ ou mais, $k \approx 0$.
- A curva cruza o zero, em média, a cada ~3ℓ.

Otimização Bayesiana

- Depois de medir: ficam só as funções que passam pelas medidas

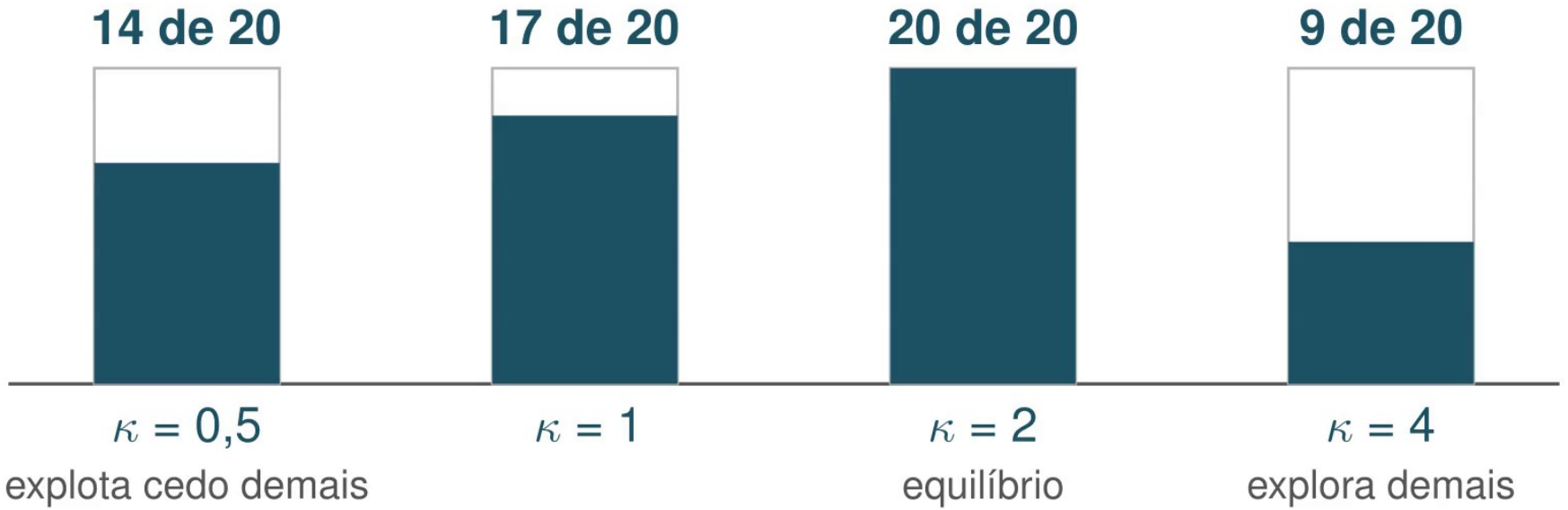


- μ(x): a média dessas funções — o palpite.
- σ(x): o quanto elas se espalham — a dúvida.
- Antes de medir, σ(x) = σ em todo lugar.
- Depois, encolhe perto das medidas.

Onde medir agora?

10. O placar do κ

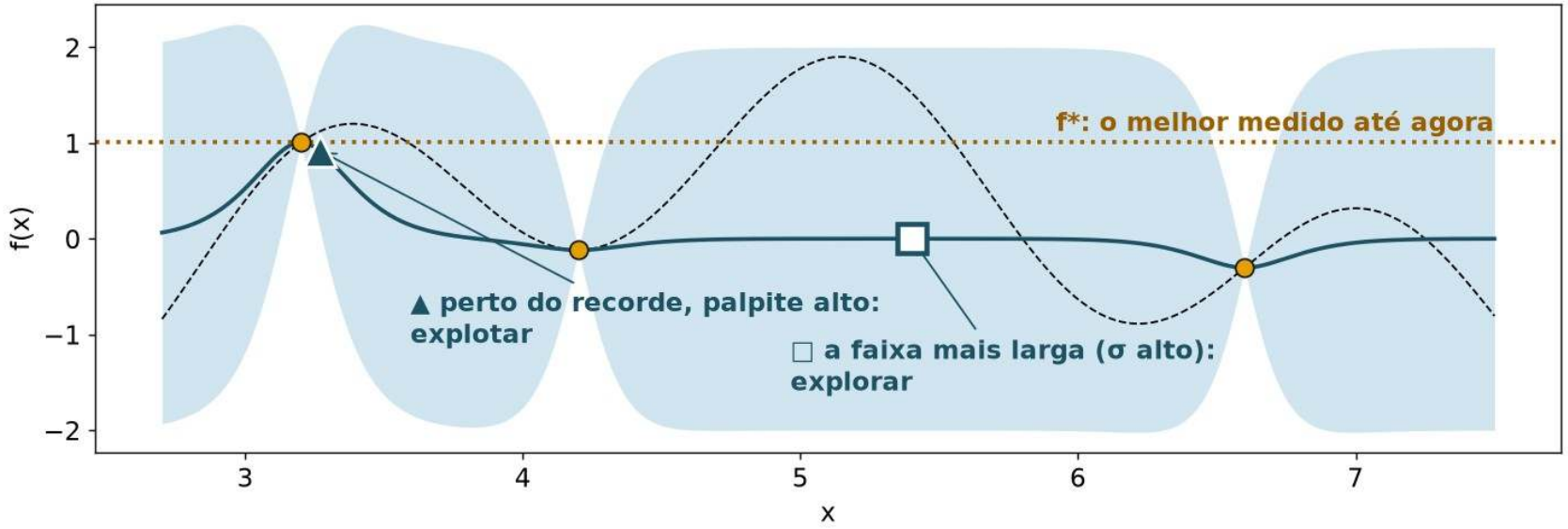
UCB, 10 medidas novas por repetição



O melhor κ depende do orçamento.

Otimização Bayesiana

- Onde medir agora? Explorar ou explotar



- Objetivo: achar o maior f(x).
- f*: o recorde a bater — o melhor valor medido até agora.
- Explotar (exploit): medir onde o palpite μ é alto.
- Explorar (explore): medir onde a dúvida σ é grande.

Otimização Bayesiana

- As três regras de aquisição
  - UCB (limite superior de confiança): o topo da faixa.

$$\mathrm{UCB} = \mu + \kappa\,\sigma$$

  - PI (probabilidade de melhora): a chance de passar de f*.

$$\mathrm{PI} = \Phi(z), \qquad z = \frac{\mu - f^*}{\sigma}$$

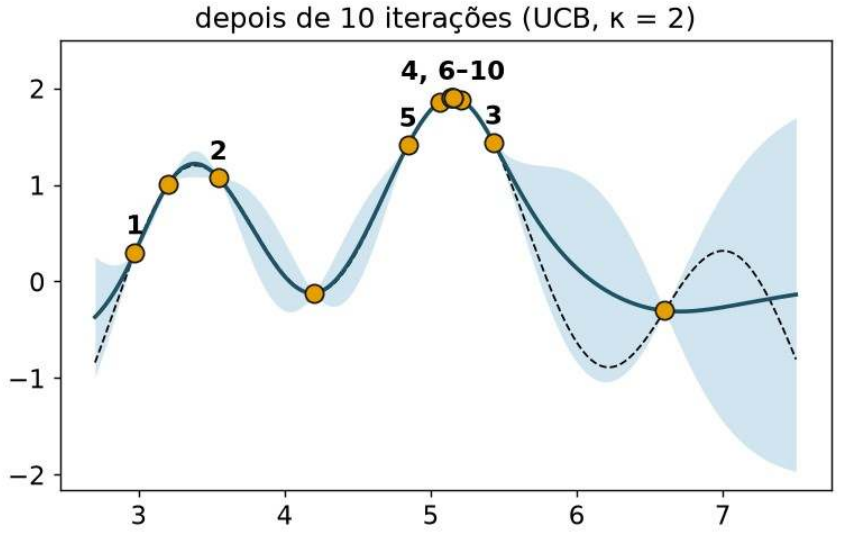
  - EI (melhora esperada): cada melhora vezes a sua chance.

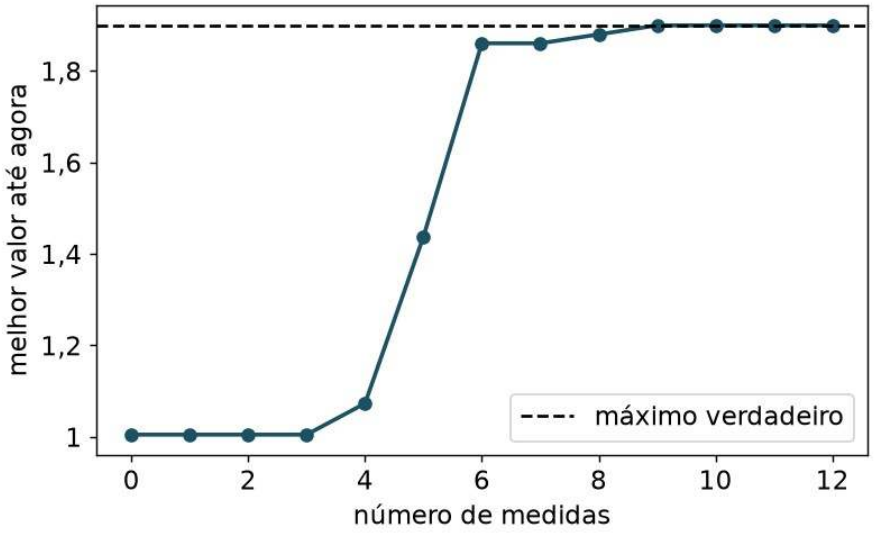
$$\mathrm{EI} = (\mu - f^*)\,\Phi(z) + \sigma\,\varphi(z)$$

  - κ: um número que nós escolhemos.
  - Φ: a área da normal padrão até z.
  - φ: a altura da normal padrão em z.

Otimização Bayesiana

- O laço da otimização bayesiana

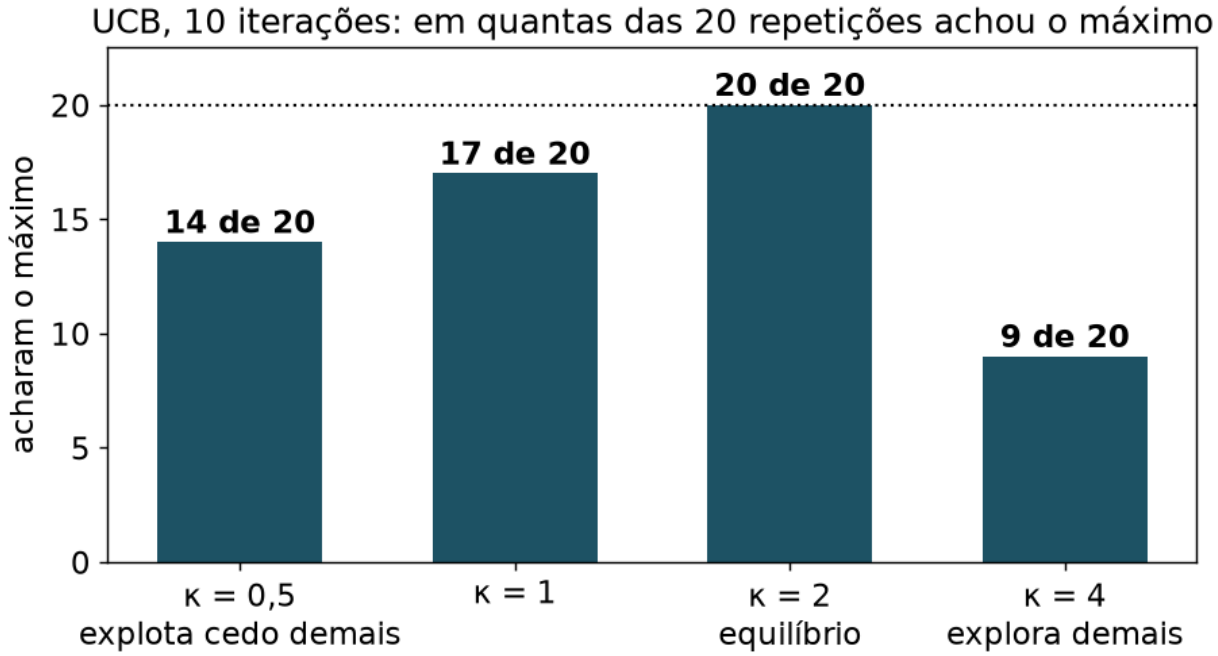




- Medidas iniciais → ajusta o GP às medidas.
- Mede onde a aquisição é máxima → repete.
- EI: achou o máximo em 13 de 20 repetições.
- UCB (κ = 2): achou em 20 de 20.

Otimização Bayesiana

- O κ: em quantas das 20 repetições achou o máximo



- κ pequeno: explota cedo demais.
- κ grande: explora demais e não refina o pico.
- O melhor κ depende do orçamento.

Otimização Bayesiana

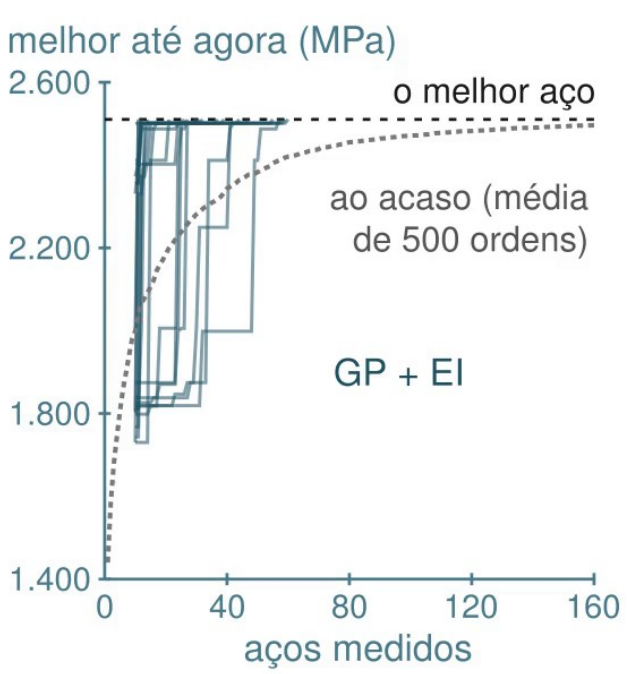
- Na prática: o Optuna faz o laço por você

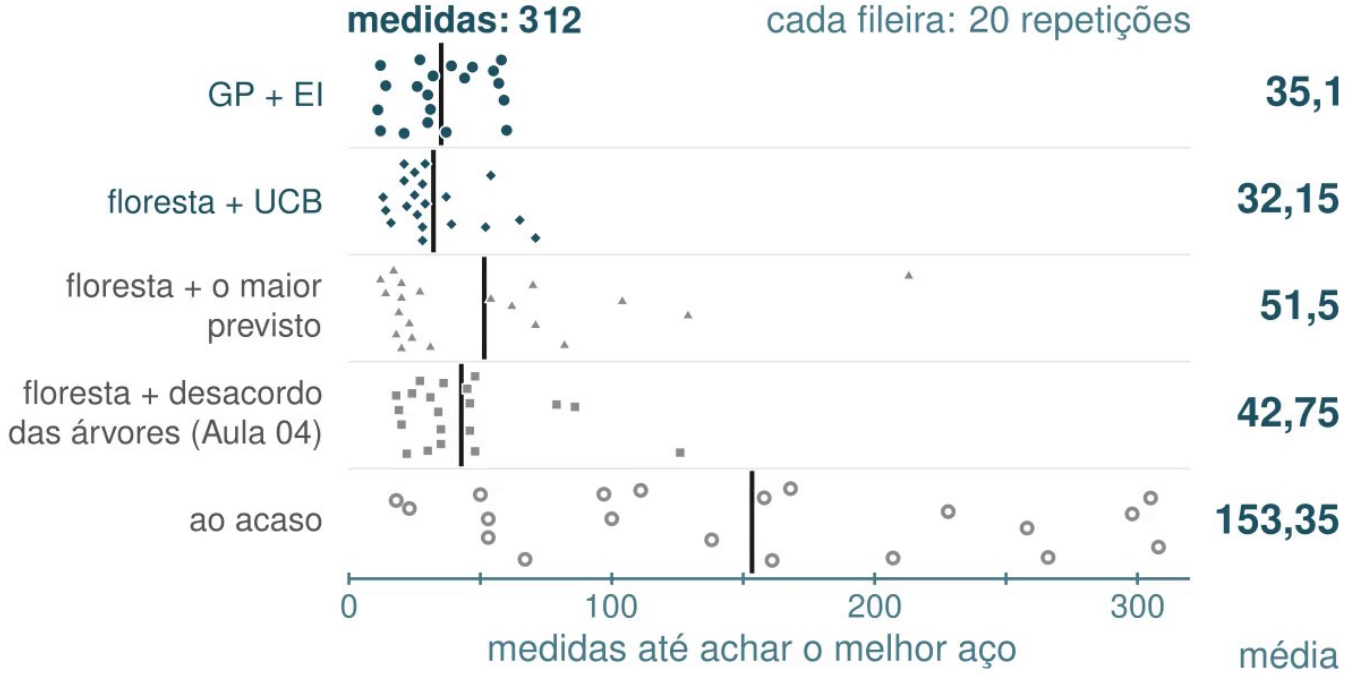
```
def objetivo(trial):
    x = trial.suggest_float("x", 2.7, 7.5)
    return f(x)            # cada chamada: uma medida cara
amostrador = optuna.samplers.GPSampler(n_startup_trials=3)
estudo = optuna.create_study(direction="maximize", sampler=amostrador)
estudo.optimize(objetivo, n_trials=13)
```

  - GPSampler: GP + EI, como o laço desta aula.
    - Sem n_startup_trials=3, as 10 primeiras são ao acaso.
  - TPESampler (Tree-structured Parzen Estimator), o padrão:
    - propõe pontos parecidos com as boas tentativas.
    - Funciona com variáveis contínuas e discretas.
  - Em 20 repetições na função da aula, achou o máximo:
    - GP em 20, TPE em 2, aleatório em 1.

Otimização Bayesiana

O enigma: 312 aços — quantas medidas até o mais resistente?

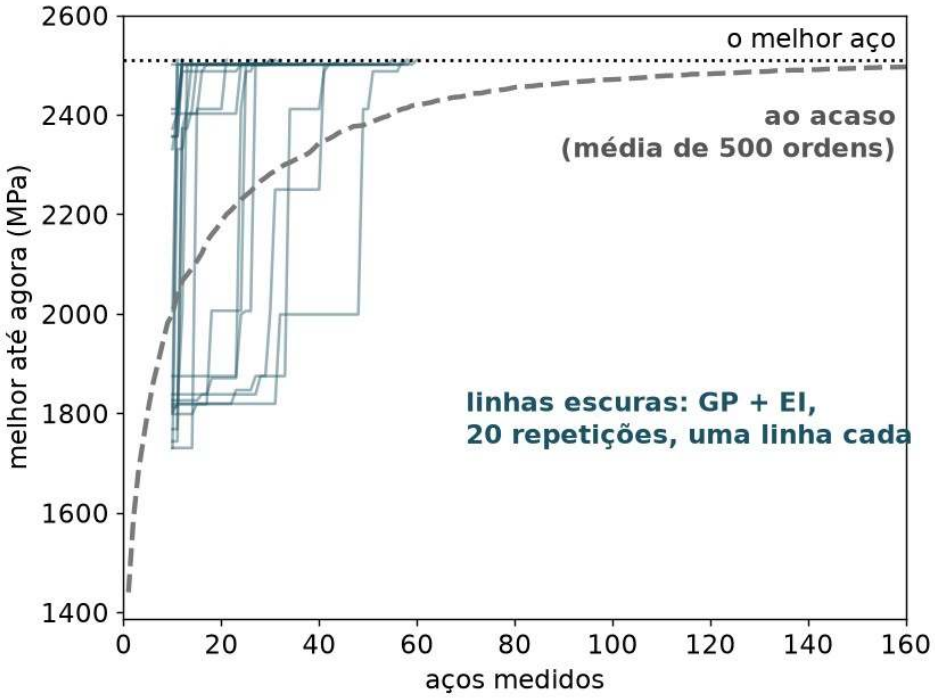


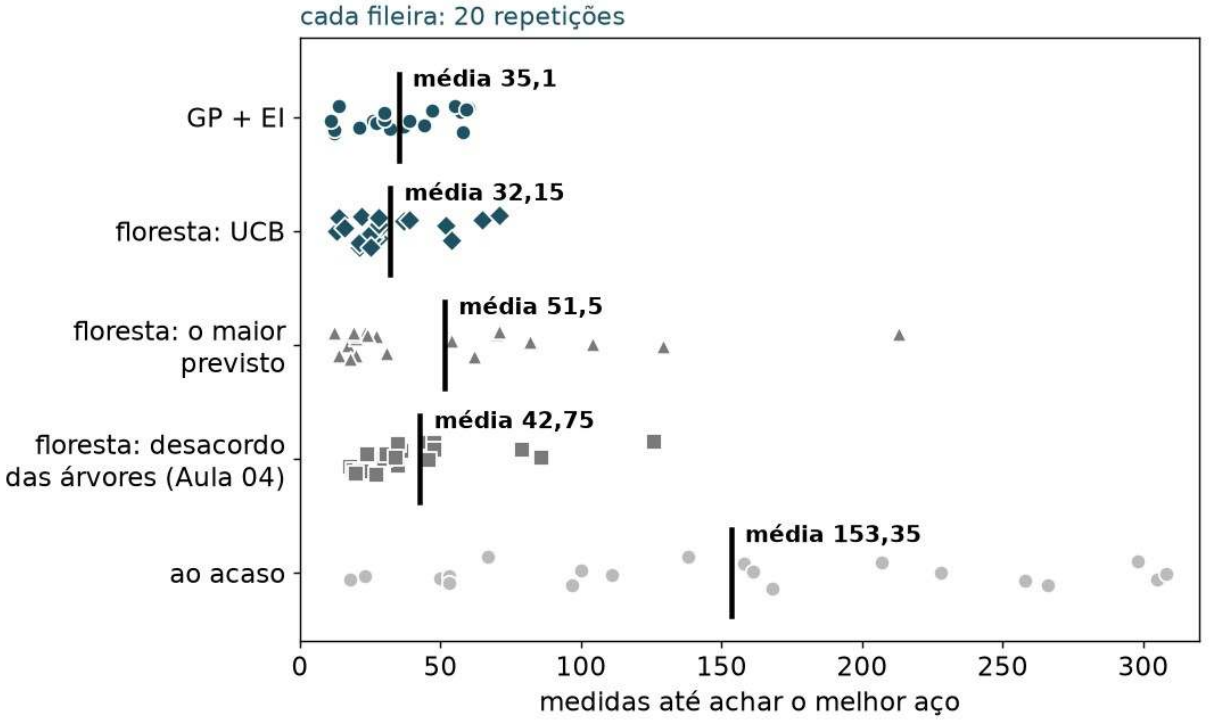


A regra pesa mais que o modelo.

Otimização Bayesiana

- O enigma: 312 aços — quantas medidas até o mais resistente?

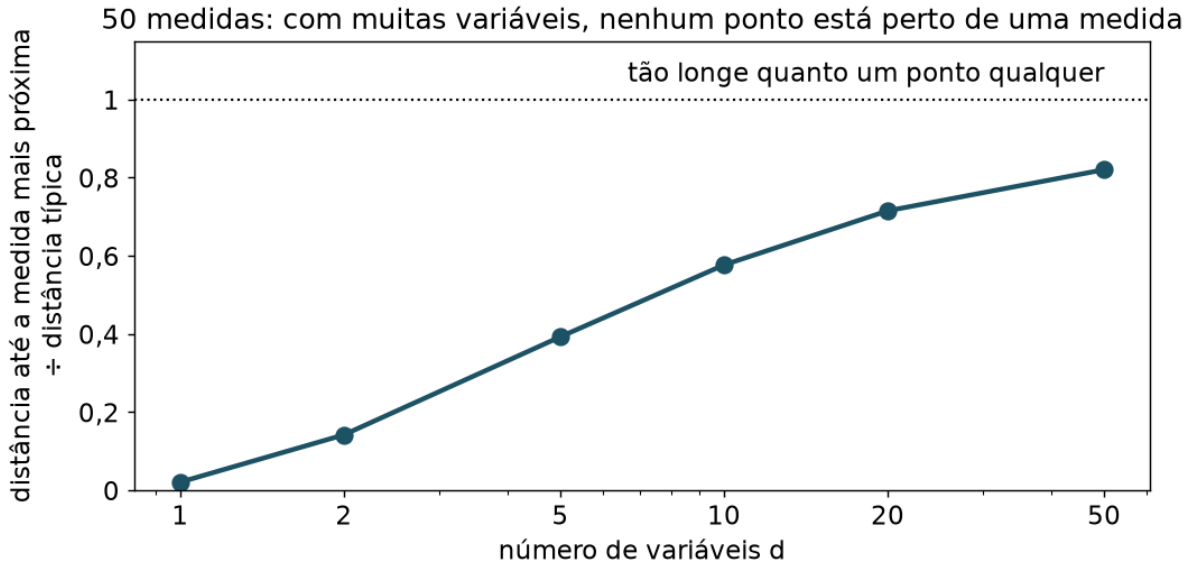




- Uma repetição: 10 aços sorteados, depois cada regra escolhe.
- 20 repetições, cada uma com outros 10 iniciais.
- O acaso também com 20: 20 ordens sorteadas.

Otimização Bayesiana

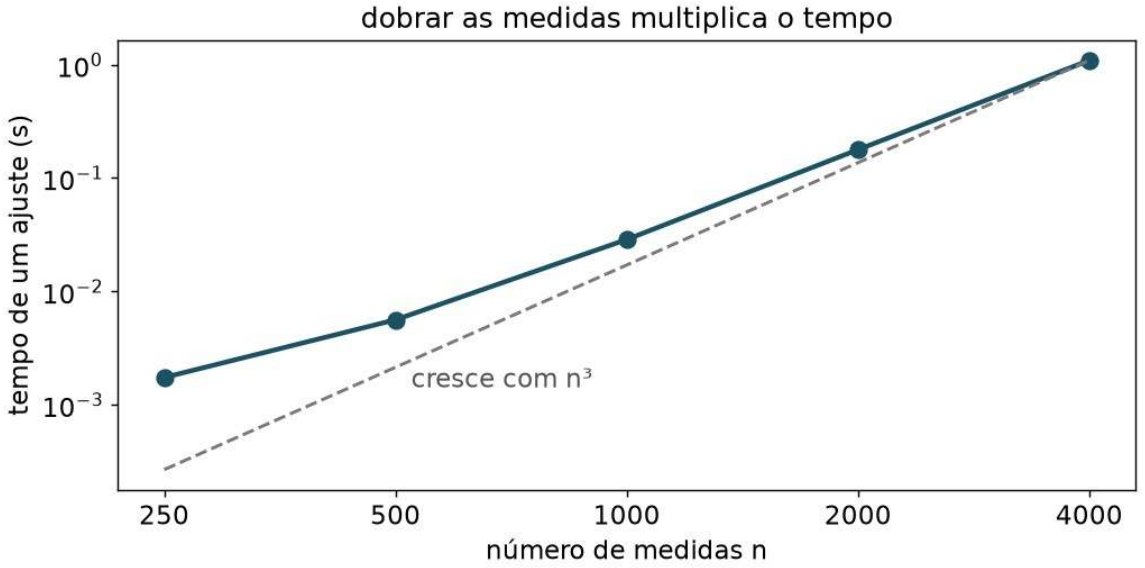
- Limite 1 — muitas variáveis: tudo fica longe de tudo



- Variáveis: as entradas do GP — o que se escolhe ao medir.
  - Na função da aula, 1. No enigma, 13 (os elementos).
- O GP decide pela distância às medidas.
- Com muitas variáveis, nenhum ponto fica perto de uma medida.
- Saída: escolher as variáveis que importam.

Otimização Bayesiana

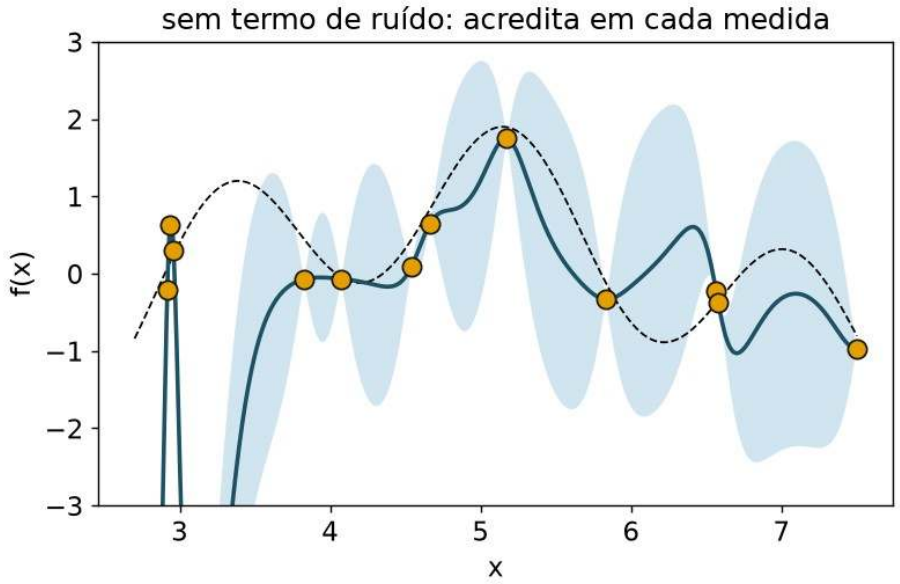
- Limite 2 — muitas medidas: o custo de ajustar o GP explode

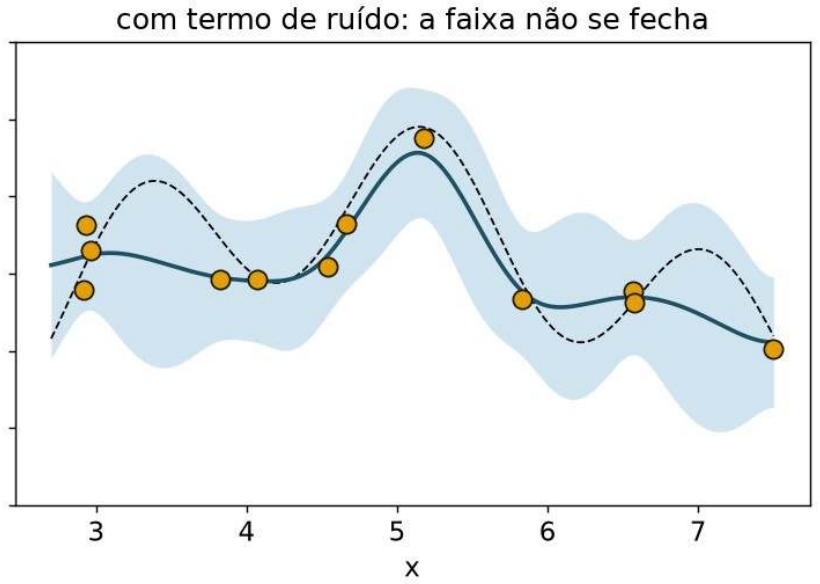


- Para μ e σ, o GP resolve n equações (n = número de medidas).
- Custo ~n^3: dobrar n → 8 vezes mais contas.
- Medido aqui, com um núcleo: × 6,1 de 2.000 para 4.000 medidas.
- Bom para 100 medidas caras, ruim para 100 mil baratas.

Otimização Bayesiana

- Limite 3 — medidas com ruído





  - Sem termo de ruído, o GP acredita em cada medida.
    - A média serpenteia.
    - A faixa se fecha onde não devia.
  - Com o termo, o GP estima o erro: 0,31, contra 0,30 de verdade.

Otimização Bayesiana

- Limite 4 — variáveis inteiras e categóricas
  - Inteiras: têm ordem e distância.
    - Número de árvores da floresta (20 a 300).
    - Número de lags da previsão (Aula 03).
    - O GP do Optuna as trata arredondando.
  - Categóricas: não têm ordem.
    - Catalisador A, B ou C.
    - Usar o calendário ou não (Aula 03).
    - Kernel RBF ou kernel Matern.
    - Para o kernel, só "igual" ou "diferente".
  - O TPE trata os dois tipos naturalmente.

Otimização Bayesiana

- Limite 5 — restrições
  - "A maior resistência com alongamento acima de 10%."
    - Alongamento: o quanto o aço estica antes de romper.
    - Abaixo de ~10%, o aço é frágil.
  - Nos nossos aços, o mais resistente (2.510,3 MPa) estica só 7,7%.
    - Com a restrição, o melhor tem 1.870,6 MPa.
  - Um segundo modelo estima a chance de cumprir a restrição.
    - Essa chance multiplica a aquisição.

Otimização Bayesiana

- A lição do bloco
  - A otimização bayesiana otimiza o número que você dá a ela.
  - Escolher esse número é o seu trabalho:
    - qual métrica, em quantas janelas, com quantas repetições.
  - É onde mora quase todo erro.

Otimização Bayesiana

- Apresentações de 19/11 — duplas, 10 minutos cada
  - Um problema do bloco, à escolha:
    - séries temporais, aprendizado ativo ou otimização bayesiana.
  - O problema e a pergunta.
  - O que fizeram e com que referência compararam.
  - O resultado medido, com várias repetições.

Otimização Bayesiana

- Vamos para o Jupyter!

OBRIGADO

James Moraes de Almeida

ilum.cnpem.br

CNPEM
Centro Nacional de Pesquisa
em Energia e Materiais