

GYRA+

Otimização Bayesiana

Sumário

Otimização	3
Parâmetros x Hiperparâmetros	6
Ajuste de Hiperparâmetros	9
Otimização Bayesiana	13

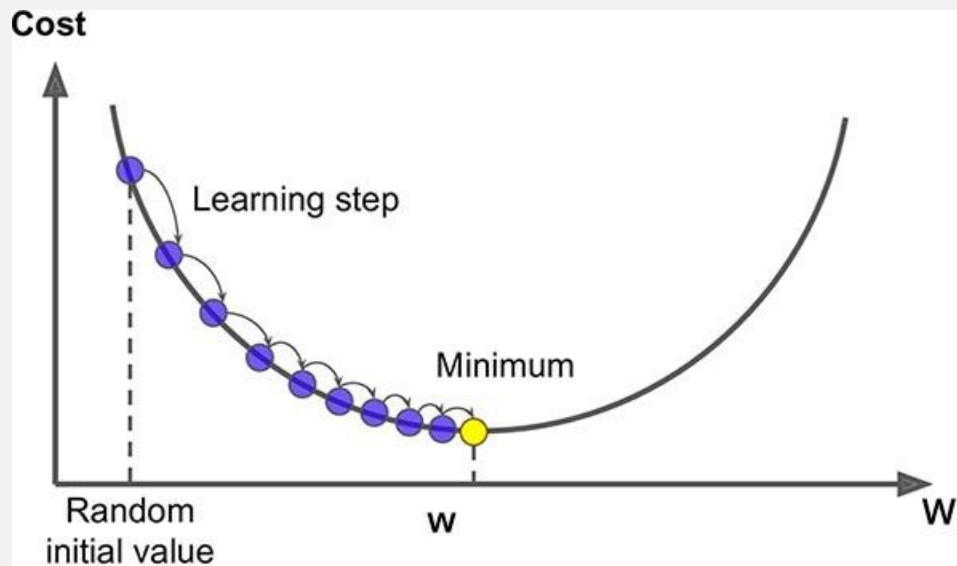
Otimização

O que é?

- Maximização ou Minimização de uma função objetivo
 - Geralmente há restrições (condições) sobre os valores das variáveis dependentes e da função
 - Solução analítica nem sempre é possível.
 - **Exemplos de Algoritmos:**
 - Método do Gradiente
 - Métodos dos Mínimos Quadrados
 - Algoritmo Simplex
-

Otimização no contexto de Machine Learning

- Treino de modelos (ajuste de parâmetros)
- Ajuste de hiperparâmetros



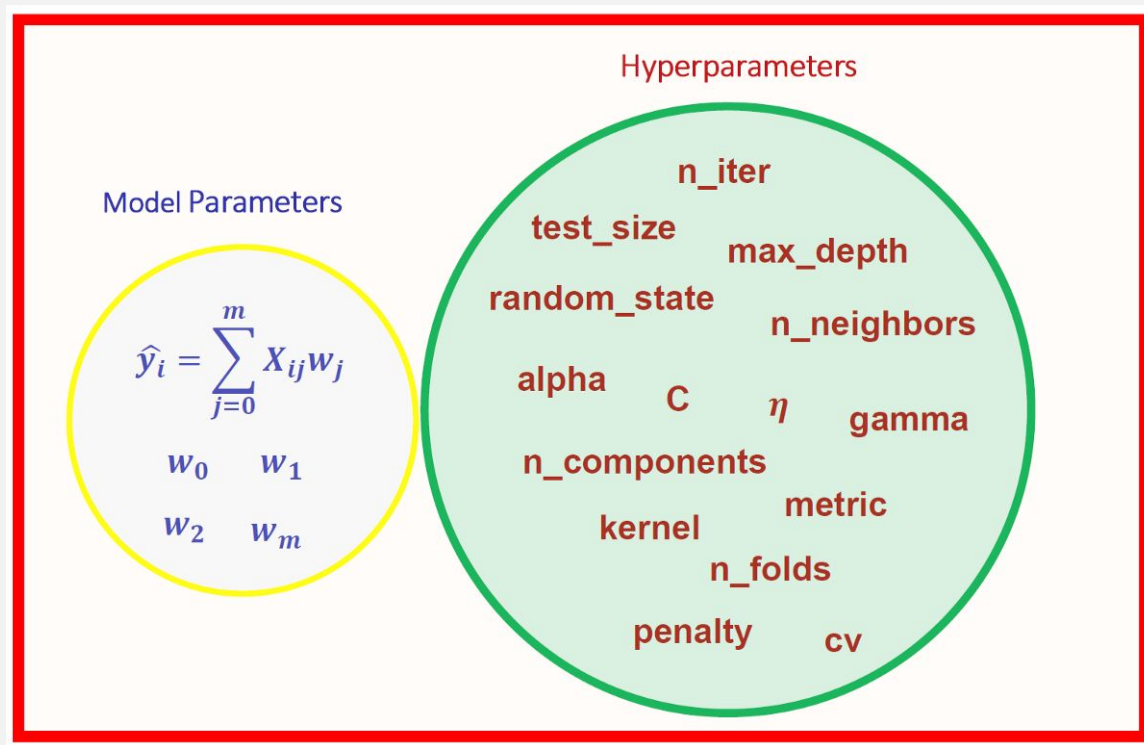
Parâmetros x Hiperparâmetros

Parâmetros

- Estimados durante o treino do modelo
- Dependem dos dados utilizados para o treino
- Exemplos: coeficientes dos vetores de suporte do SVM, coeficientes da Regressão Linear.

Hiperparâmetros

- Definidos antes do treino
 - Independentes dos dados de treino
 - Exemplos: número de vizinhos do K-Neighbours, número de camadas de uma rede neural.
-



Ajuste de Hiperparâmetros

Otimização de Hiperparâmetros

- Melhora no valor da métrica alvo
- Evita problemas de super e sub ajuste dos dados de treino.
- Pode ser resumido como:

$$\theta^* = \arg \min_{\theta \in \Theta} H(\theta)$$

- Onde

$$H : \Theta \rightarrow \mathbb{R}$$

Input (hyper-parameter space Θ , Target score function $H(\theta)$, max n° of evaluation n_{max})

Select an initial hyper-parameter configuration $\theta_0 \in \Theta$

Evaluate the initial score $y_0 = H(\theta_0)$

Set $\theta^ = \theta_0$ and $y^* = H(\theta_0)$*

For $n=2, \dots, n_{max}$ **do**

Select a new hyper-parameter configuration $\theta_n \in \Theta$ using some optimization strategy

Evaluate H to obtain a new numeric score $y_n = H(\theta_n)$

If $y_n < y^*$

$\theta^ = \theta_n$ and $y^* = y_n$*

End if

end for

Output: θ^* and y^*

Estratégias de Otimização

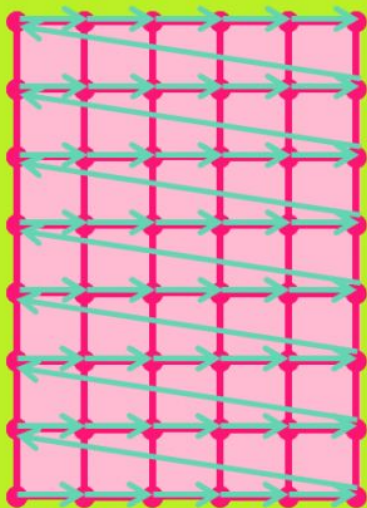
Não Bayesiana

- Cada varredura ocorre de forma independente.
- Exemplos:
 - Grid Search
 - Random Search
- Função objetivo tem forma desconhecida

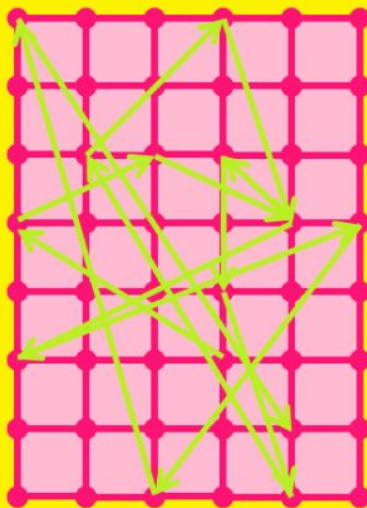
Bayesiana

- Baseada no Teorema de Bayes
 - Nova varredura ocorre com informação de varreduras passadas
 - Função objetivo é a realização de um processo estocástico
-

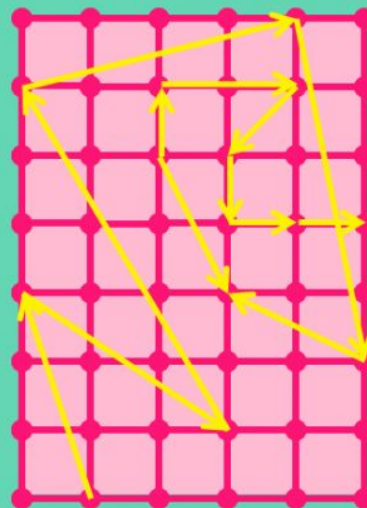
GRID



RANDOM



BAYESIAN



Otimização Bayesiana

Idéia Básica

- **Teorema de Bayes**

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

- **Função Surrogate:** distribuição a priori da função objetivo.
- **Função de Aquisição:** função utilizada para definir a nova amostragem dos hiperparâmetros

Input (hyper-parameter space Θ , Target score function $H(\theta)$, max n° of evaluation n_{max})

Select an initial configuration $\theta_0 \in \Theta$

Evaluate the initial score $y_0 = H(\theta_0)$

Set $\theta^* = \theta_0$, $y^* = H(\theta_0)$, and $S_0 = \{\theta_0, y_0\}$

For $n=1, \dots, n_{max}$ **do**

Select a new hyper-parameter configuration $\theta_n \in \Theta$ by optimizing an acquisition function U_n

$$\theta_n = \arg \max_{\theta \in \Theta} U_n(\theta; S_t),$$

Evaluate H in θ_n to obtain a new numeric score $y_n = H(\theta_n)$

Augment the data $S_n = S_{n-1} \cup \{\theta_n, y_n\}$

Update the surrogate model

If $y_n < F^*$

$\theta^* = \theta_n$ and $y^* = y_n$

End if

end for

Output: θ^* and y^*

Função Surrogate

Modelagem com Processo Gaussiano

$$H(\theta) \sim GP(\mu(\theta); k(\theta, \theta'))$$

$$P(y|\theta, S) = GP(\mu(\theta)_{H|S}, k(\theta, \theta')_{H|S})$$

Input (hyper-parameter space Θ , Target score function $H(\theta)$, max n° of evaluation n_{max})

Select an initial configuration $\theta_0 \in \Theta$

Evaluate the initial score $y_0 = H(\theta_0)$

Set $\theta^* = \theta_0$, $y^* = H(\theta_0)$, and $S_0 = \{\theta_0, y_0\}$

For $n=1, \dots, n_{max}$ **do**

Select a new hyper-parameter configuration $\theta_n \in \Theta$ by optimizing an acquisition function U_n

$$\theta_n = \arg \max_{\theta \in \Theta} U_n(\theta; S_t),$$

Evaluate H in θ_n to obtain a new numeric score $y_n = H(\theta_n)$

Augment the data $S_n = S_{n-1} \cup \{\theta_n, y_n\}$

Update the surrogate model

If $y_n < F^*$

$\theta^* = \theta_n$ and $y^* = y_n$

End if

end for

Output: θ^* and y^*

Funções de Aquisição

- Probability of Improvement (PI):

$$PI(\theta) = \int_{-\infty}^{y^*} P(y|\theta, S) dy$$

- Expected Improvement (EI):

$$EI(\theta) = \int_{-\infty}^{y^*} (y - y^*) P(y|\theta, S) dy$$

Input (hyper-parameter space Θ , Target score function $H(\theta)$, max n° of evaluation n_{max})

Select an initial configuration $\theta_0 \in \Theta$

Evaluate the initial score $y_0 = H(\theta_0)$

Set $\theta^* = \theta_0$, $y^* = H(\theta_0)$, and $S_0 = \{\theta_0, y_0\}$

For $n=1, \dots, n_{max}$ **do**

Select a new hyper-parameter configuration $\theta_n \in \Theta$ by optimizing an acquisition function U_n

$$\theta_n = \arg \max_{\theta \in \Theta} U_n(\theta; S_t),$$

Evaluate H in θ_n to obtain a new numeric score $y_n = H(\theta_n)$

Augment the data $S_n = S_{n-1} \cup \{\theta_n, y_n\}$

Update the surrogate model

If $y_n < F^*$

$\theta^* = \theta_n$ and $y^* = y_n$

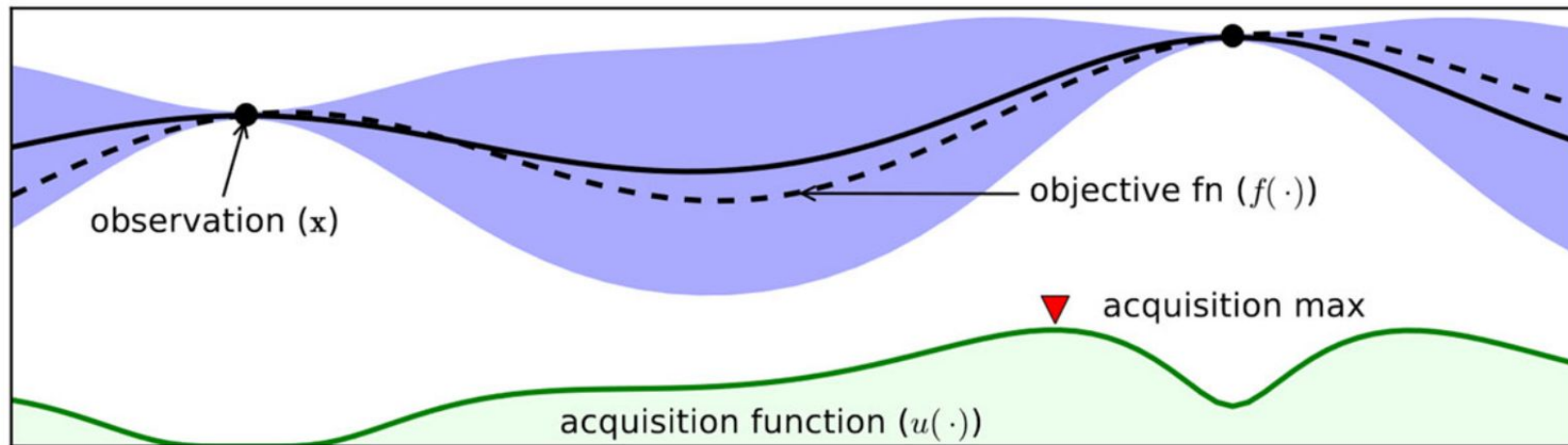
End if

end for

Output: θ^* and y^*

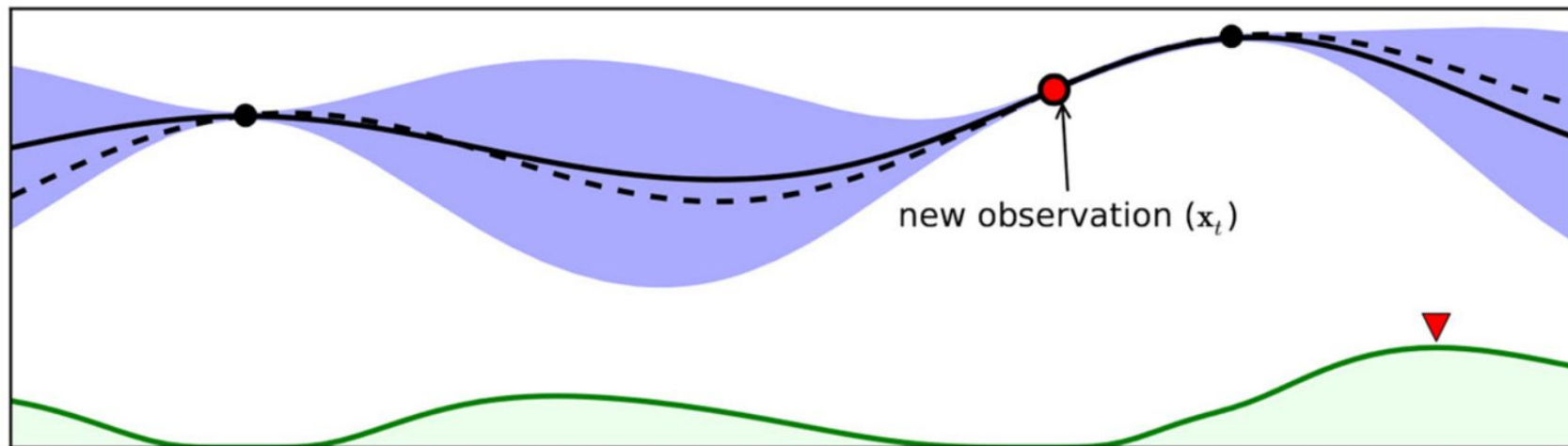
Visualização

$t = 2$



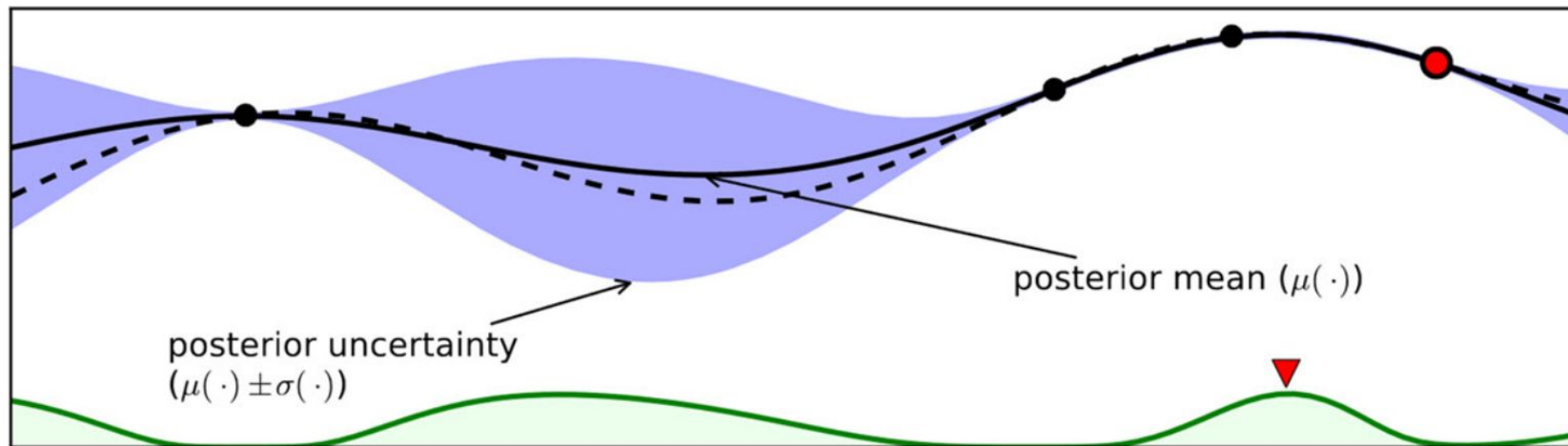
Visualização

$t = 3$



Visualização

$t = 4$



Desvantagens da Técnica

- Sensibilidade aos parâmetros da função surrogate
 - Número de amostras necessárias para aperfeiçoar a função surrogate aumenta com o aumento do espaço de hiperparâmetros
 - Custo computacional pode ser alto em relação a Random Search
-

Bibliotecas

- Scikit-optimize:
https://scikit-optimize.github.io/stable/auto_examples/bayesian-optimization.html#bayesian-optimization-loop
 - HyperOpt: <http://hyperopt.github.io/hyperopt/#getting-started>
-

Links

- Processo Gaussiano Iterativo:
<https://distill.pub/2019/visual-exploration-gaussian-processes/>
 - Bayesian Optimization is Superior to Random Search for Machine Learning Hyperparameter Tuning: <https://arxiv.org/abs/2104.10201>
 - Random Search for Hyper-Parameter Optimization, Bergstra & Bengio 2012:
<https://www.jmlr.org/papers/volume13/bergstra12a/bergstra12a.pdf>
 - Aulas: <https://www.cs.ubc.ca/~nando/540-2013/lectures.html>
-

Obrigado/a.

Rua Farne de Amoedo 76, Sala 403

Ipanema - Rio de Janeiro
