

---

# Sessão 06

WiDS Recife Live Coding, 25/01/2020

---

# Como serão os live codings?

- Sessões ao-vivo todos os sábados das 14h às 15h
  - Código e slides serão disponibilizados no nosso site
  - O objetivo é treinar para participar do [Datathon](#) em 2020
-

---

# Roteiro

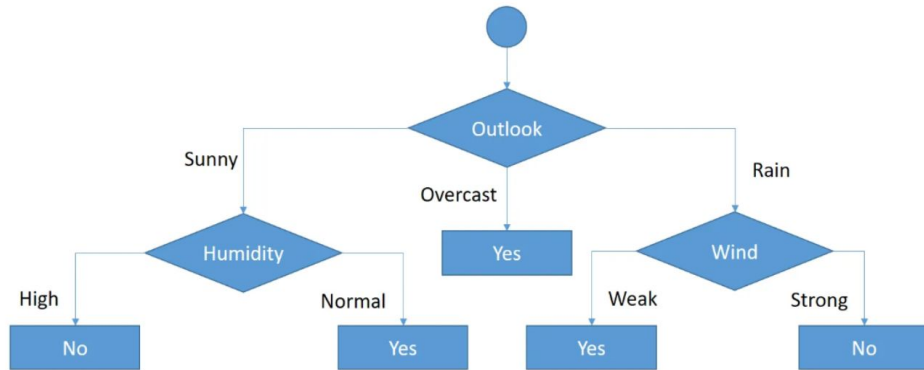
- Entendimento do algoritmo do Xgboost.
- Hiper-parâmetros
- Apresentação de código

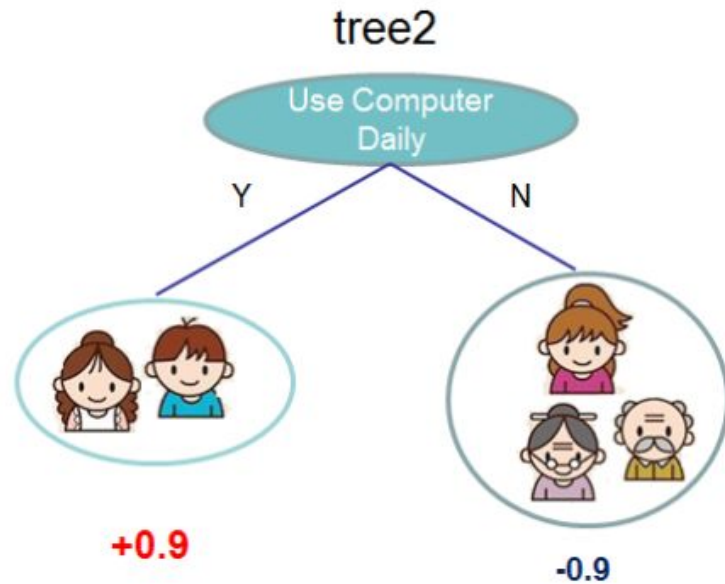
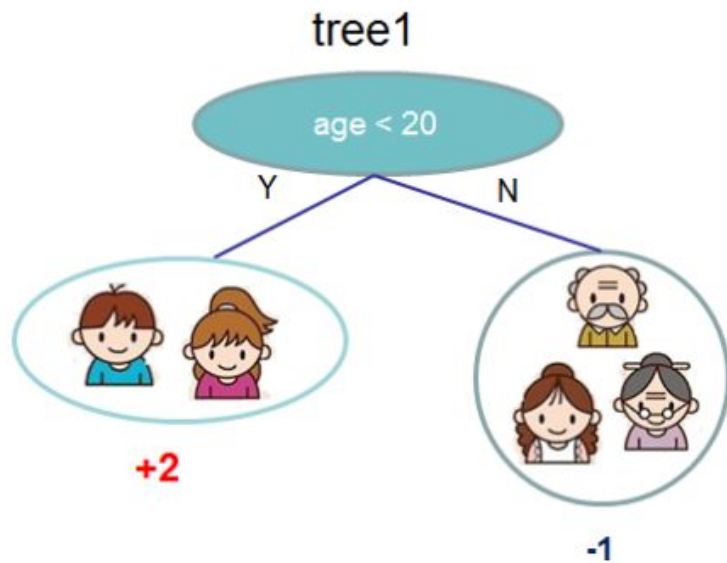
# Xgboost

- O Xgboost é um algoritmo de aprendizado de máquina utilizado em problemas supervisionados, como por exemplo regressão ou classificação.

Árvores de Decisão são métodos onde existe uma função que recebe um vetor de valores (de atributos) como entrada e retorna uma decisão (de saída).

# Árvores de Decisão



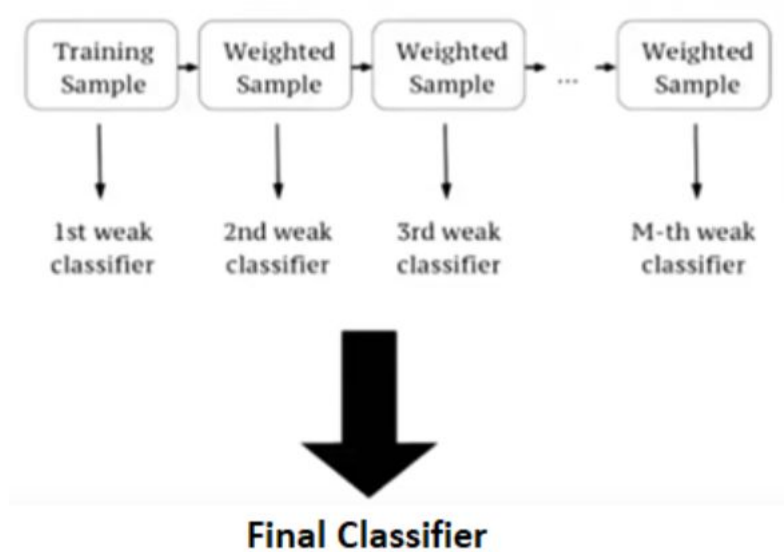


$f(\text{young person}) = 2 + 0.9 = 2.9$

$f(\text{old person}) = -1 - 0.9 = -1.9$

# Gradient Boosting Machine

- Xgboost é baseado no GBM.
- O método do gradiente (ou método do máximo declive) é um método numérico usado em otimização.
- Combina resultados de muitos classificadores “fracos”, tipicamente árvores de decisão, para formar algo parecido com um “comitê forte de decisão”



- No xGboost a regra mais fraca é definida através de uma árvore de decisão.



# Vantagens do uso do xgb

- Aderente a uma ampla variedade de aplicações;
- Portabilidade;
- Integração em nuvem

# Hiper-parâmetros

# Parâmetros de otimização

Parâmetros que podem ser aperfeiçoados e ajustados para se adequar melhor a um problema específico. Quando bem definidos, são capazes de aumentar consideravelmente a performance de um modelo.

---

- 
- **max\_depth** - Define a profundidade máxima da árvore;
  - **Min\_child\_weight** - Soma mínima dos pesos das instâncias necessárias em um filho;
  - **Subsample** - Proporção de subamostras do treinamento;
  - **Colsample\_bytree** - Proporção de subamostras de colunas ao construir a árvore;
  - **nround** - Número de árvores
-

---

# Obrigada!

E até semana que vem!

---

# —

## Referências

- [1] <https://xgboost.readthedocs.io/en/latest/parameter.html>
- [2] <https://xgboost.readthedocs.io/en/latest/tutorials/model.html>
- [3] <https://www.datacamp.com/community/tutorials/xgboost-in-python>