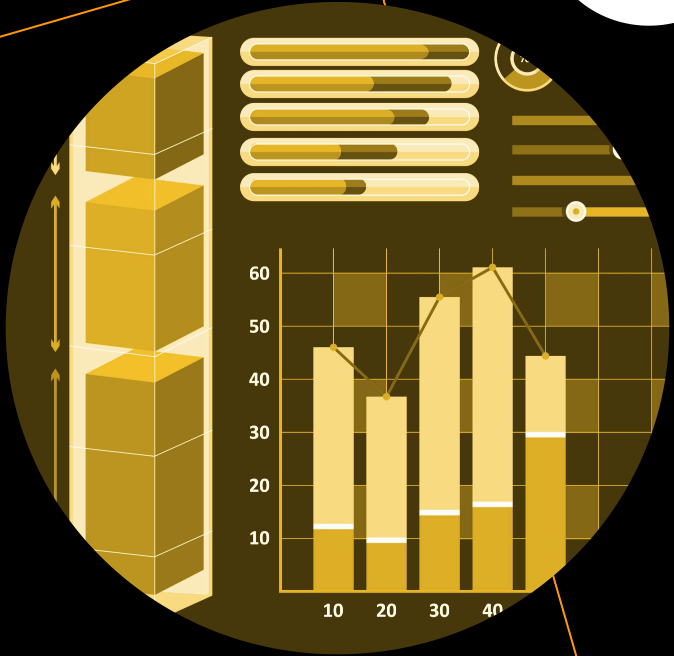
MaSSP

Bagging - Random Forest

Nhung Le



Giới thiệu về ensemble methods

Ensembles - Parallel vs. Sequential

- **Ensembles:** Sử dụng nhiều models
- **Parallel ensembles:**
  - Goal: Develop models in parallel to reduce variance
  - Bagging (e.g., Random Forests)
- **Sequential ensembles:**
  - Goal: Develop models sequentially so the next one learns from the previous one
  - Boosting (e.g., Gradient Boosting)

Bootstrap

Bootstrapping method

A statistical method to estimate quantities about a population by averaging estimates from multiple small data samples from the original population.

Steps:

- Suppose we have 1000 data points → D = 1000
- We get 15 datasets, each has 700 randomly chosen data points from our pool of 1000 data points → B = 15, n = 700
- Learning algorithm gives B decision functions:

$$\hat{f}_1(x), \hat{f}_2(x), \ldots, \hat{f}_B(x)$$

Bagging

Bagging

Bagging = Bootstrap Averaging

Bootstrapping Steps:

- Suppose we have 1000 data points → D = 1000
- We get 15 datasets, each has 700 randomly chosen data points from our pool of 1000 data points → B = 15, n = 700
- Learning algorithm gives B decision functions:

$$\widehat{f_1}(x), \widehat{f_2}(x), \ldots, \widehat{f_B}(x)$$

Bagging

Bagging = Bootstrap Averaging

Averaging Steps:

- We fix some particular x_0
- Then we have

$$\widehat{f}_{avg}(x_0) = \frac{1}{B}\sum_{b=1}^{B}\widehat{f}_b(x_0)$$

Bagging

MaSSP

Expected value

$$
\begin{aligned}
\mathbb{E}(\widehat{f}_{avg}(x_0)) &= \mathbb{E}\left(\frac{1}{B}\sum_{b=1}^{B}\widehat{f}_b(x_0)\right) \\
&= \frac{1}{B}\mathbb{E}\left(\sum_{b=1}^{B}\widehat{f}_b(x_0)\right) \\
&= \frac{1}{B}\sum_{b=1}^{B}\mathbb{E}(\widehat{f}_b(x_0)) \\
&= \frac{1}{B}B.\mathbb{E}(\widehat{f}_i(x_0)) \qquad \forall i = 0, 1, ..., B \\
&= \mathbb{E}(\widehat{f}_i(x_0)) \qquad \forall i = 0, 1, ..., B
\end{aligned}
$$

Conclusion:

$\hat{f}_{avg}(x_0)$ and $\hat{f}_b(x_0)$ have the same expected value since $\hat{f}_1(x_0), \hat{f}_2(x_0), ..., \hat{f}_b(x_0)$ are independent

Bagging

Variance

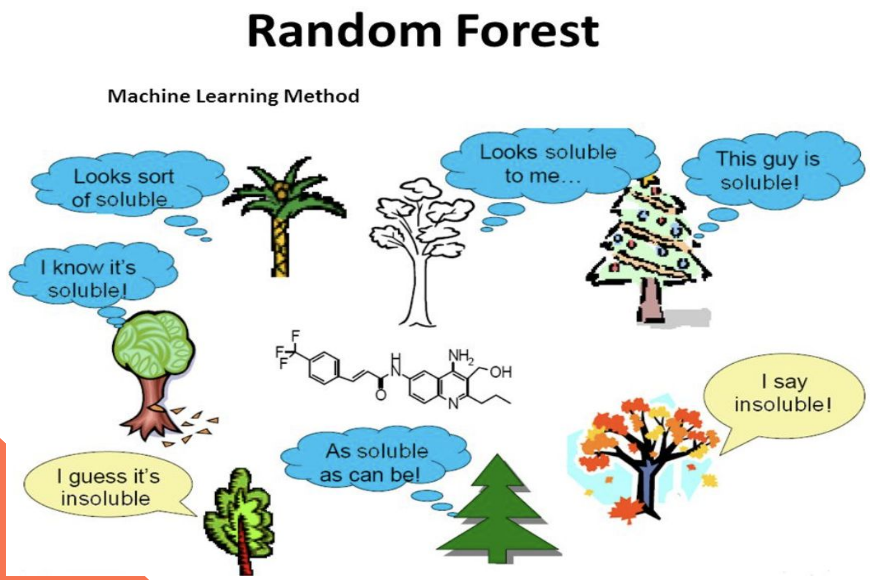
$$\begin{aligned}
Var(\widehat{f}_{avg}(x_0)) &= Var\left(\frac{1}{B}\sum_{b=1}^{B}\widehat{f}_b(x_0)\right) \\
&= \frac{1}{B^2} Var\left(\sum_{b=1}^{B}\widehat{f}_b(x_0)\right) \\
&= \frac{1}{B^2} B.Var(\widehat{f}_i(x_0)) \qquad \forall i = 0, 1, ..., B \\
&= \frac{1}{B} Var(\widehat{f}_i(x_0)) \qquad \forall i = 0, 1, ..., B.
\end{aligned}$$

Conclusion: $\hat{f}_{avg}(x_0)$ has smaller variance than $\hat{f}_b(x_0)$ note that $\hat{f}_1(x_0), \hat{f}_2(x_0), ..., \hat{f}_b(x_0)$ are independent

Random Forests

Random Forest

Use bagged decision tree and grow each tree independently to reduce dependence between trees.



Trong rừng cây, mỗi cây sẽ đưa ra một quyết định.

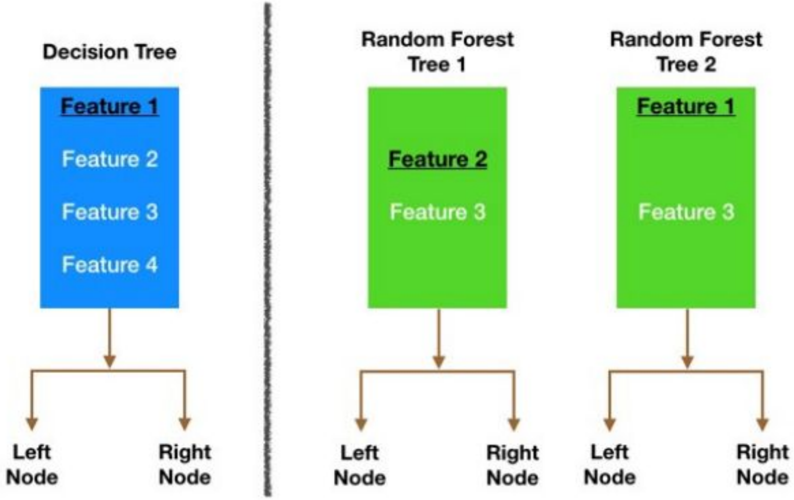
Kết quả cuối cùng có thể tính bằng cách **1/ lấy giá trị trung bình** hoặc **2/ quyết định của số đông.**

Random Forest

- When growing each tree, restrict choice of spliting variables to a randomly chosen subset of features of size m
- We often choose $m = \sqrt{p}$ where p is the number of features

Random Forest

MaSSP



Node splitting in a random forest model is based on a random subset of features for each tree.

- **Random sampling:** *Mỗi cây trong Random Forest sẽ được huấn luyện (train) với **một nhóm data ngẫu nhiên***
- **Random feature selection:** *mỗi cây trong Random Forest sẽ sử dụng một **nhóm features** (subset of features) **khác nhau** để đảm bảo sự độc lập (independence)*

Nguồn: towarddatascien

MaSSP

MaSSP

Appendix