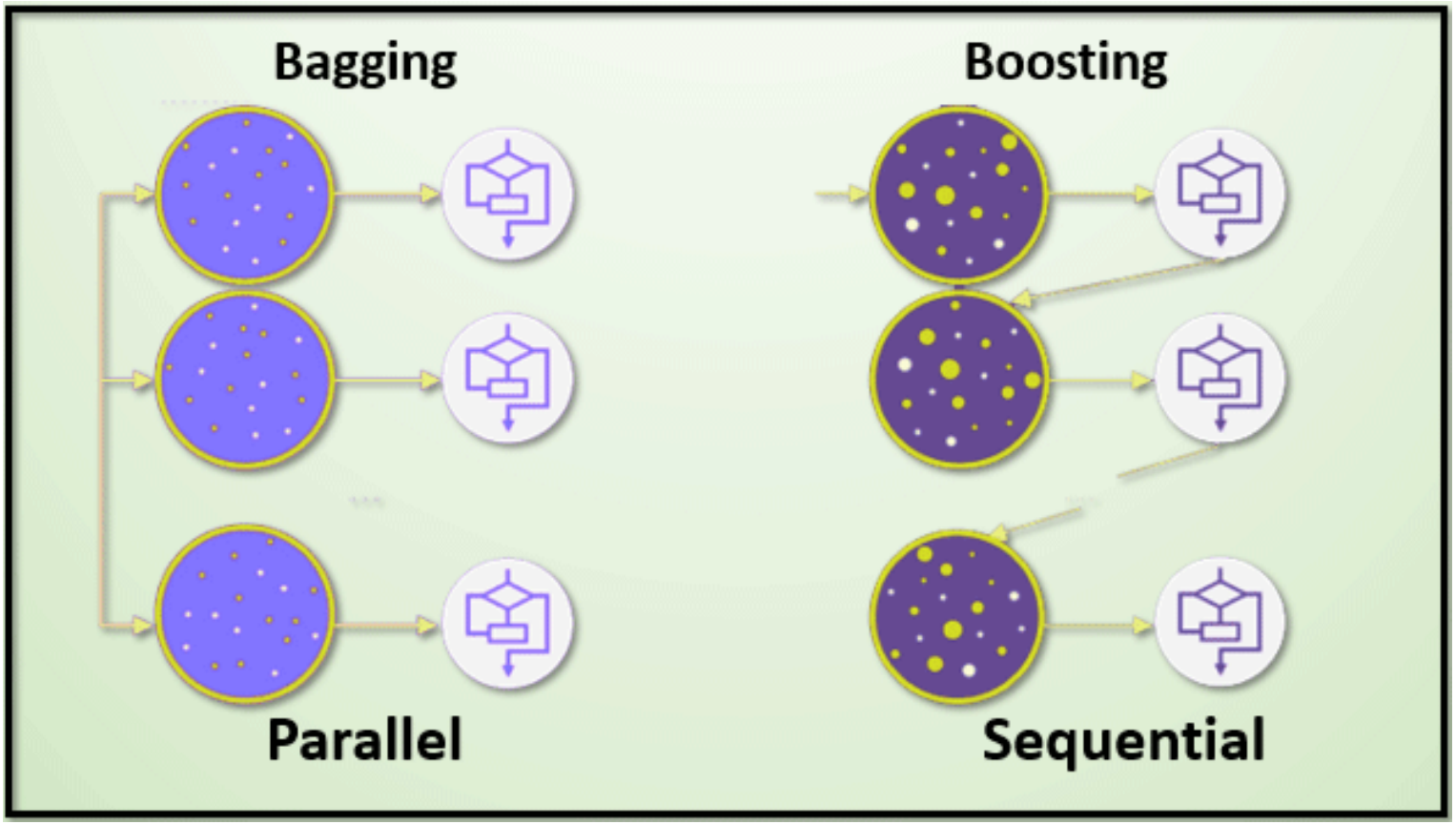

Primena i poređenje *boosting*-a i *bagging*-a

Kada i kako primeniti AdaBoost?

- AdaBoost je dizajniran za binarnu klasifikaciju
  - Probleme sa više kategorija možemo tretirati kao više binarnih klasifikacionih problema
  - Ali, preduslov za dobre performanse AdaBoost je $\varepsilon(f_t) < 0.5)$. Za binarni klasifikator, ovo znači da radi bolje od šanse. Kod višekategorijskih problema, ovo je teže postići
  - Bolja opcija je AdaBoost prilagođen da radi sa višekategorijskim klasifikacionim problemima

    [Mukherjee, I. and Schapire, R.E., 2013. A theory of multiclass boosting. *Journal of Machine Learning Research*, *14*(Feb), pp.437-497.]

- AdaBoost se može prilagoditi i za probleme regresije

  [Drucker, H., 1997, July. Improving regressors using boosting techniques. In ICML (Vol. 97, pp. 107-115)]

Kada i kako primeniti AdaBoost?

- AdaBoost se može koristiti za boosting performansi bilo kog ML algoritma
- Najbolje ga je koristiti sa slabim prediktorima
- Algoritam koji najviše odgovara za AdaBoost je stablo odlučivanja sa jednim nivoom dubine (*decision stump*)

Priprema podataka

- **Kvalitetni podaci**
  - AdaBoost se stalno prilagođava pogrešno klasifikovanim primerima, zbog toga moramo da pazimo da je kvalitet podataka visok
  - Šum u podacima, pogotovo u ciljnoj varijabli je veoma problematičan

- **Outlieri**
  - Dobra strana: Boosting može da identifikuje outliere jer se fokusira na primere teške za klasifikaciju
  - Međutim, previše outliera može značajno da redukuje performanse i dramatično poveća vreme potrebno za konvergenciju

Varijante boosting algoritama

- Postoji stotine varijanti boosting algoritma, *AdaBoost* je jedan od najranijih

- Danas najčešće korišćen je *gradient boosting*
  - Sličan *AdaBoost* algoritmu, ali nešto kompleksniji
  - Može se koristiti ne samo za klasifikaciju, već i za druge tipove problema

Uticaj boostinga na ML

- Ogroman – među najkorisnijim ML metodama koje su kreirane
  - Jednostavan pristup
  - Može se primeniti na bilo koji model

- *Computer vision* – standardni pristup za prepoznavanje lica
- Korišćen od strane najvećeg broja pobednika ML takmičenja (*Kaggle*, *KDD cup*,...)
  - *Malware classification, credit fraud detection, ads click through rate estimation, sales forecasting, ranking webpages for search, Higgs boson detection,...*
- Široko korišćen u industriji – modeli primenjeni u praksi najčešće su bazirani na ansamblu klasifikatora
  - Npr. *Netflix* koristi boosting za preporuku filmova

Poređenje bagginga i boostinga

- Povezanost pojedinačnih modela
  - U baggingu se svaki model gradi nezavisno
  - U boostingu se svaki model gradi na osnovu prethodnog
- Predikcija finalnog modela
  - U baggingu su svi glasovi jednaki (*simple majority vote*)
  - U boostingu se vrši otežinjeno glasanje (*weighted majority vote*)
- Efikasnost
  - Bagging je jednostavniji, a nezavisnost modela čini da ih možemo obučavati u paraleli

Poređenje bagginga i boostinga

- Način pobojšanja generalizacije

- Bagging redukuje varijansu osnovnog modela
  - Pogodan je za modele sa velikom varijansom a malim sistematskim odstupanjem (kompleksne modele)

- Boosting može da redukuje i varijansu i sistematsko odstupanje osnovnog modela
  - U boostingu se trening greška stabilno smanjuje
  - Pogodan je za jednostavne (slabe) klasifikatore niske varijanse a velikog sistematskog odstupanja

Poređenje bagginga i boostinga

- Sa aspekta performansi nema jasnog pobednika – zavisi od podataka

- Boosting obično rezultuje većim performansama od bagginga (za isto *M*) ali postoje i drugi faktori koje treba uzeti u obzir

Poređenje bagginga i boostinga

- Sklonost overfittingu
  - Bagging rešava problem overfittinga
  - Što se tiče overfittinga, Boosting je robustniji u odnosu na korišćen osnovni model, ali sam nije imun na overfitting
  - Bagging je obično efektiviniji od boostinga ako nemamo veliko sistematsko odstupanje i samo želimo da smanjimo varijansu (npr. ako overfitujemo)

- Šum i outlieri u podacima
  - Bagging je efektivniji od Boostinga ako podaci sadrže šum ili outliere