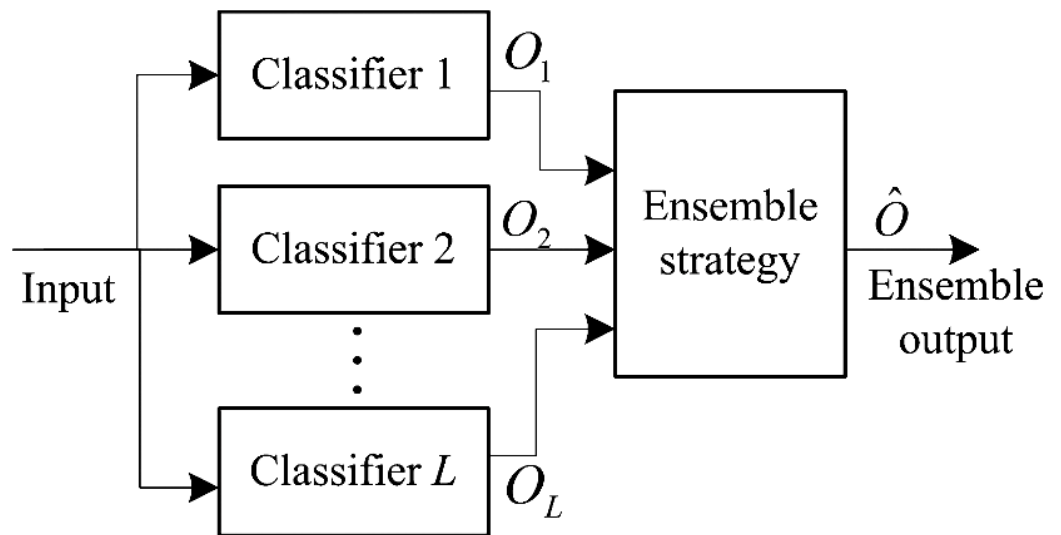


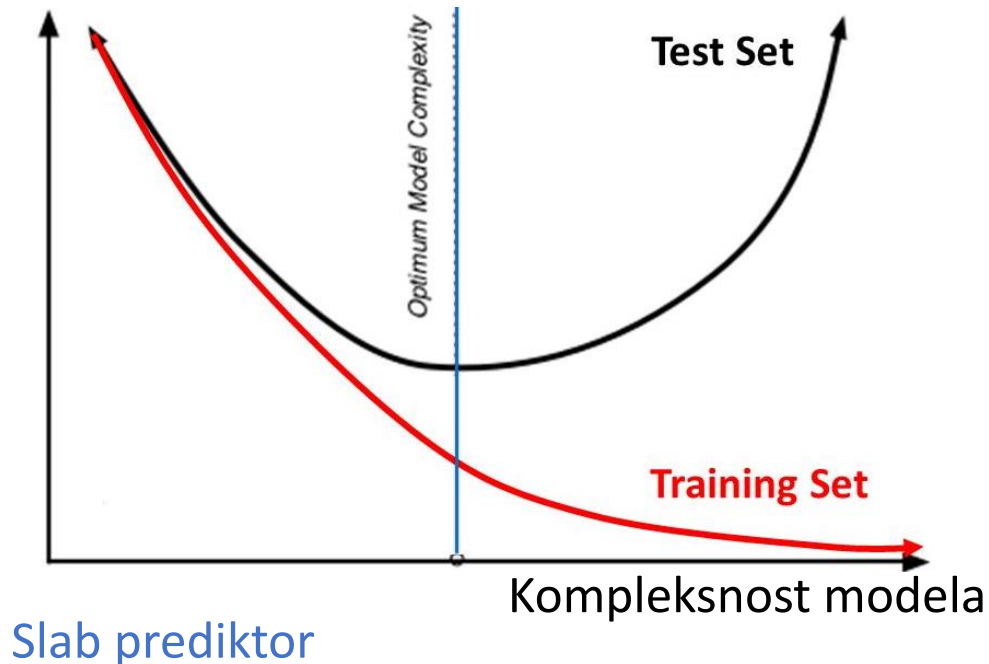
Ansambl (grupa) prediktora



Jednostavni (slabi) prediktori

- Jednostavni (slabi) prediktori

- Npr. logistička regresija sa jednostavnim obeležjima, perceptron, Naivni Bajes, plitka stabla odlučivanja,...
- Prediktor čije su performanse relativno slabe (nešto je bolji od slučajnosti ali ne mnogo)
- Računski jednostavni



Prednosti

- mala varijansa (stabilni)
- nisu skloni overfittingu

Mane

- veliko sistematsko odstupanje
- ne mogu da reše složene probleme

Kako poboljšati slab prediktor?

1. Dodati obeležja

- Npr. polinomijalna obeležja u slučaju logističke regresije i pokušati izbeći overfitting

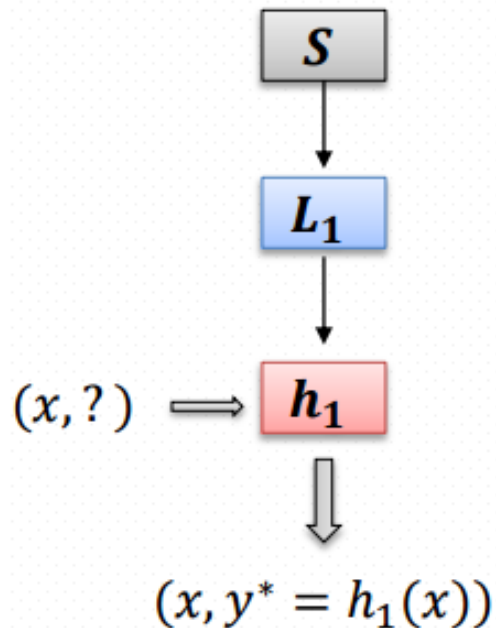
2. Kombinovanje više slabih prediktora

- Može li skup slabih klasifikatora biti kombinovan u cilju formiranja jačeg klasifikatora? *Kearns and Valiant (1988)*

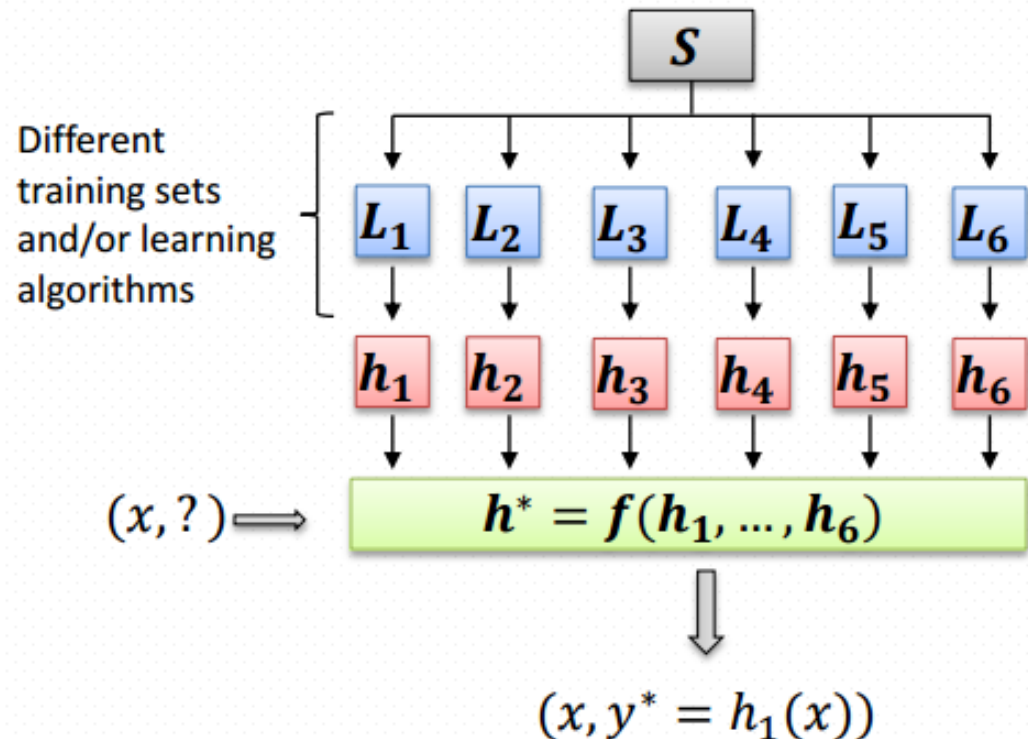
Model ansambla

- Glavna ideja:
 - obučiti više modela da rešavaju isti problem
 - kombinovati njihove glasove

Traditional:



Ensemble Method:



Kako kombinovati glasove različitih modela?

- Regresija: prosek
 - Klasifikacija: glasanje
 - *Stacking* (kasnije)
- } možemo dodeliti težine

Kako obučiti više različitih modela?

1. Manipulacija obučavajućeg algoritma

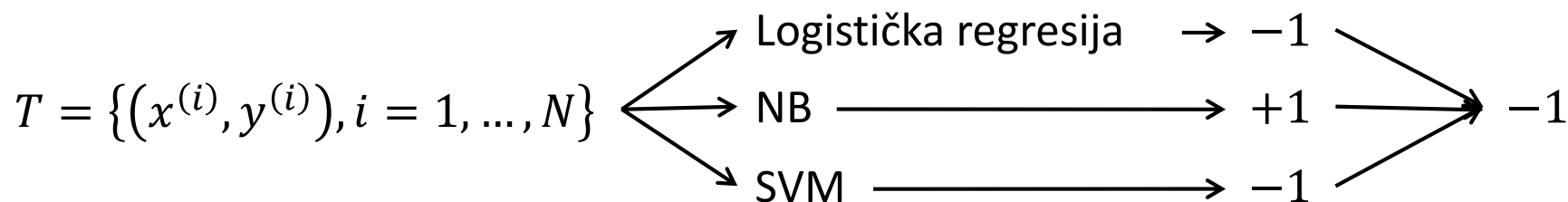
2. Manipulacija trening skupom

3. Manipulacija skupa obeležja

4. Manipulacija klasnih oznaka

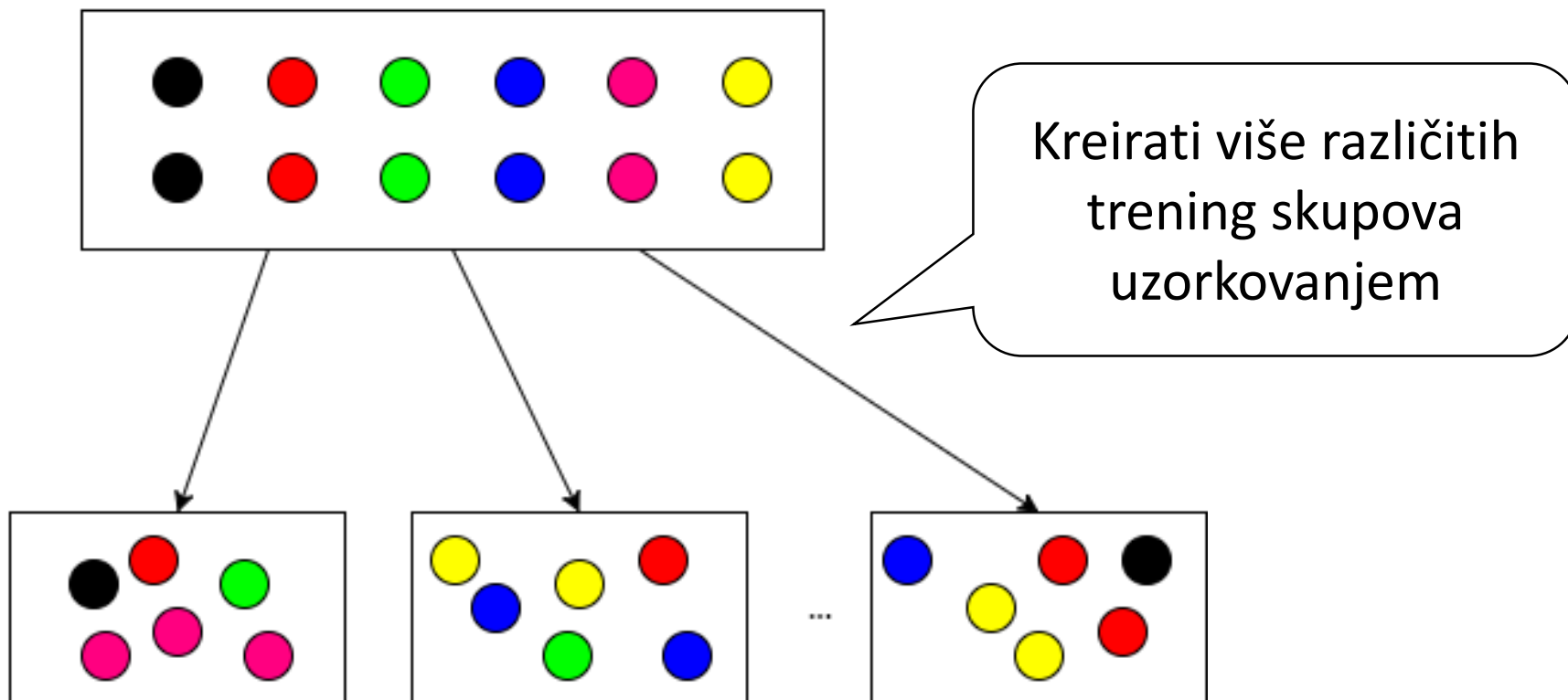
1. Manipulacija obučavajućeg algoritma

Na *istim* podacima trenirati
različite modele *nezavisno*



Možemo koristiti različite
algoritme ili isti algoritam sa
različitim postavkama parametara

2. Manipulacija trening skupom



3. Manipulacija skupa obeležja

x_1	x_2	x_3	x_4
245	26	416	5
453	67	357	15
189	47	25	45

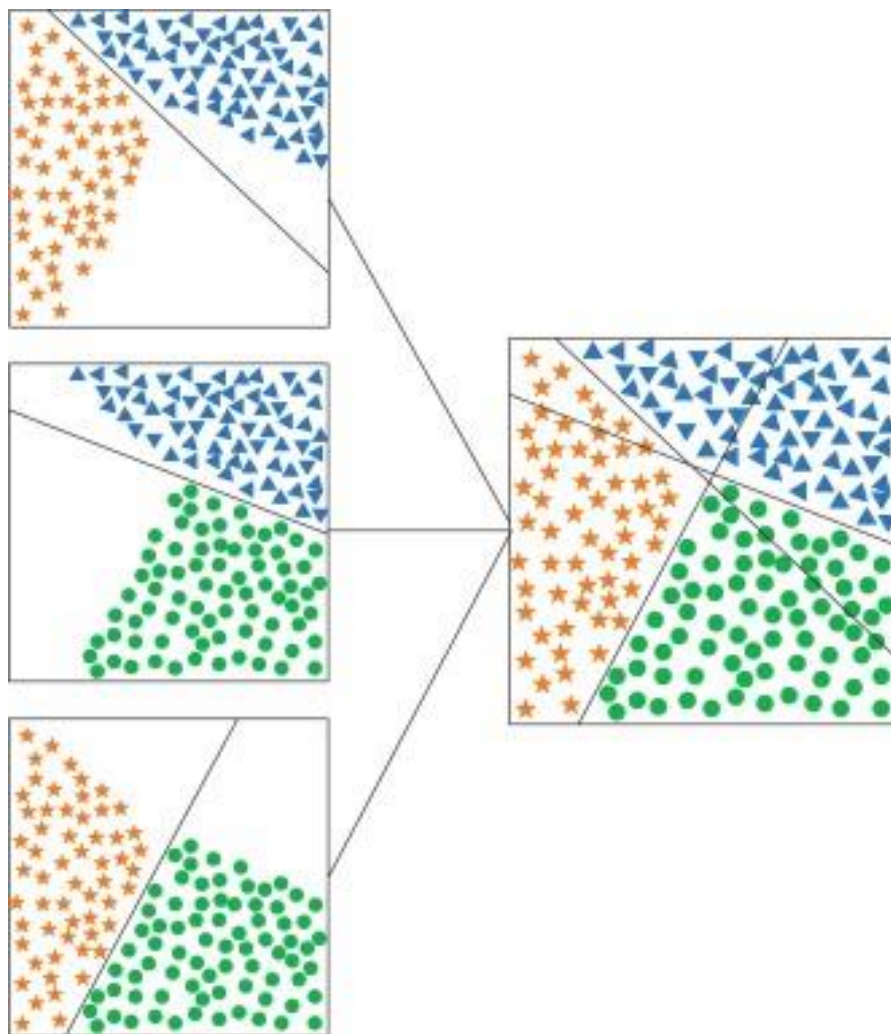
Isti podaci, različiti
podskupovi obeležja
(npr. *Random Forest*)

x_1	x_3	x_4
245	416	5
453	357	15
189	25	45

x_2	x_3	x_4
26	416	5
67	357	15
47	25	45

x_1	x_2	x_4
245	26	5
453	67	15
189	47	45

4. Manipulacija klasnih oznaka



- *One-vs-one*
- *One-vs-all*
- Veliki broj klasa
 - generisati L binarnih particija
 - trenirati L binarnih klasifikatora
 - klasifikacija: klasa čije particije dobiju najviše glasova