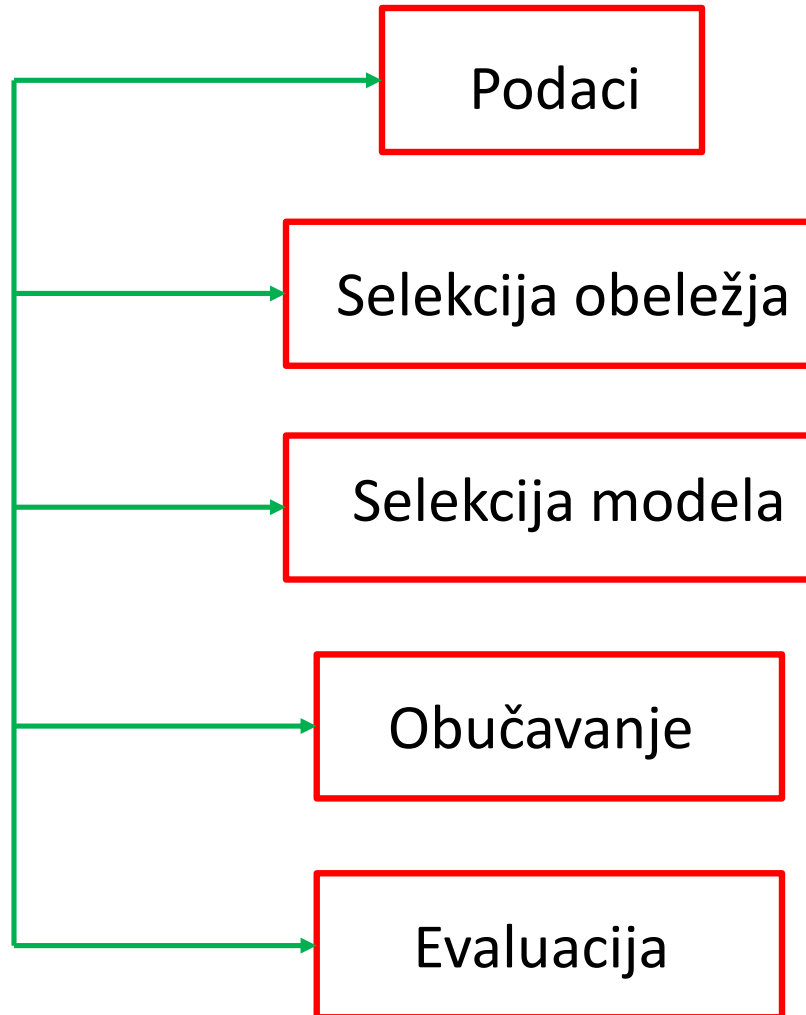


Evaluacija obučavajućih sistema

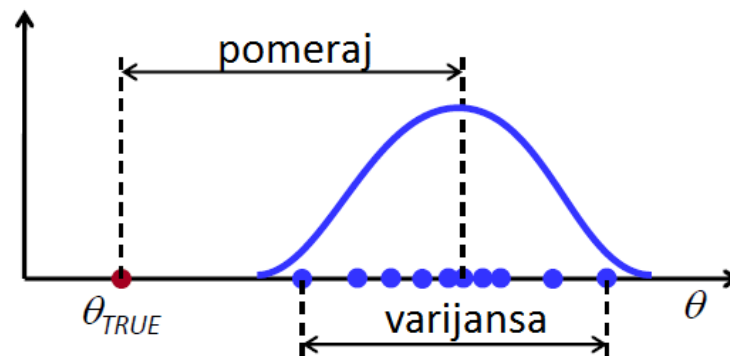
Milan M.Milosavljević

Cuklus sinteze

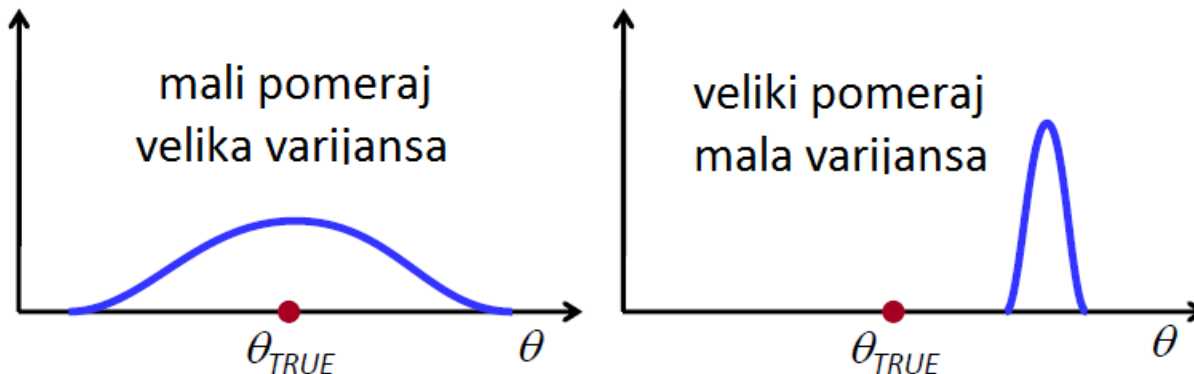


Pomeraj i varijansa

- Pomeraj (Bias) – koliko je procena blizu istinite vrednosti ?
- Varijansa – koliko se procena menja od jednog do drugog uzorka podatak ?



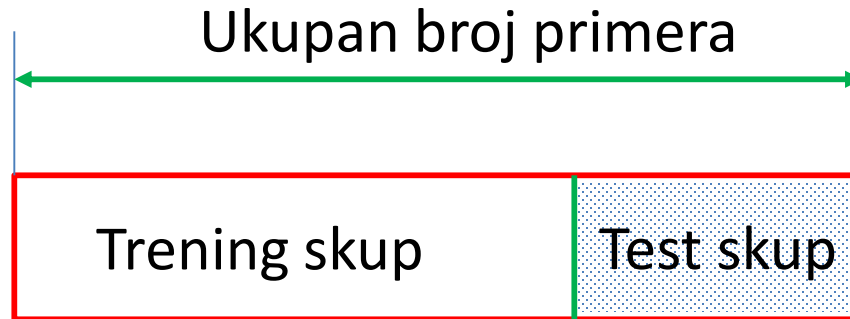
- Kompromis izmedju varijanse i pomeraja
- U većini situacija, moguće je smanjiti jednu od ovih veličina samo na račun druge



Evaluacija obučavanja

- Holdout
- Kros – validacija
 - Slučajno semplovanje
 - K- tostruka kros-validacija
 - Leave-one-out krosvalidacija
- Bustrep (Bootstrap)

Holdout metod

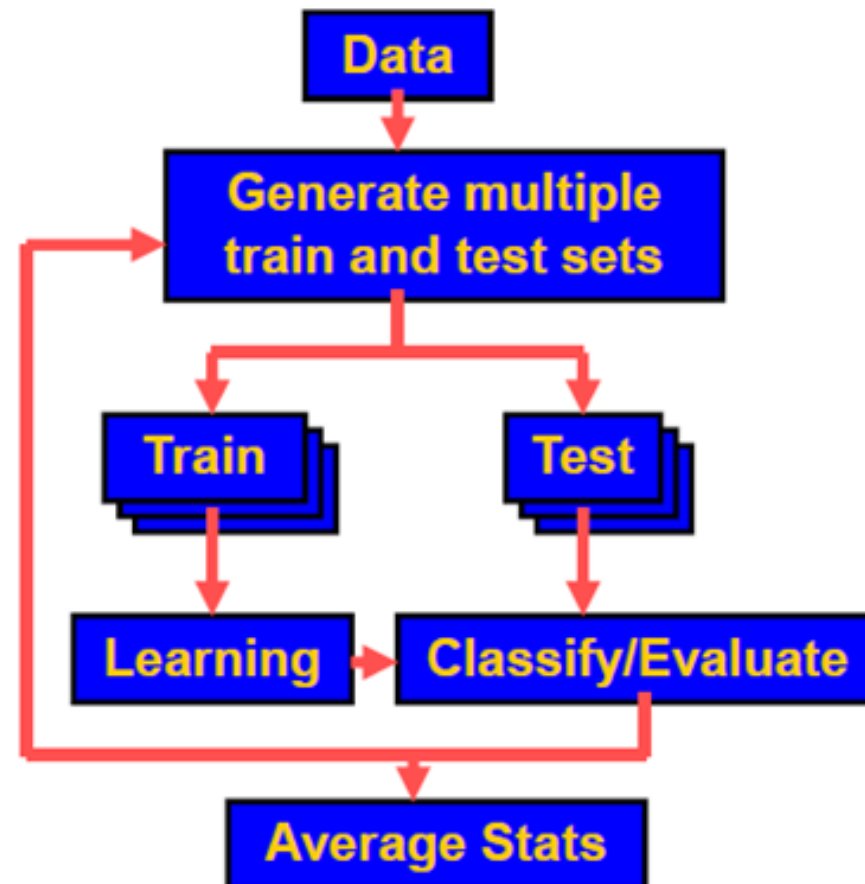


- Podeliti podatke u dve grupe:
 - Trening skup – za obuku klasifikatora
 - Test skup – za ocenu greške obučenog klasifikatora

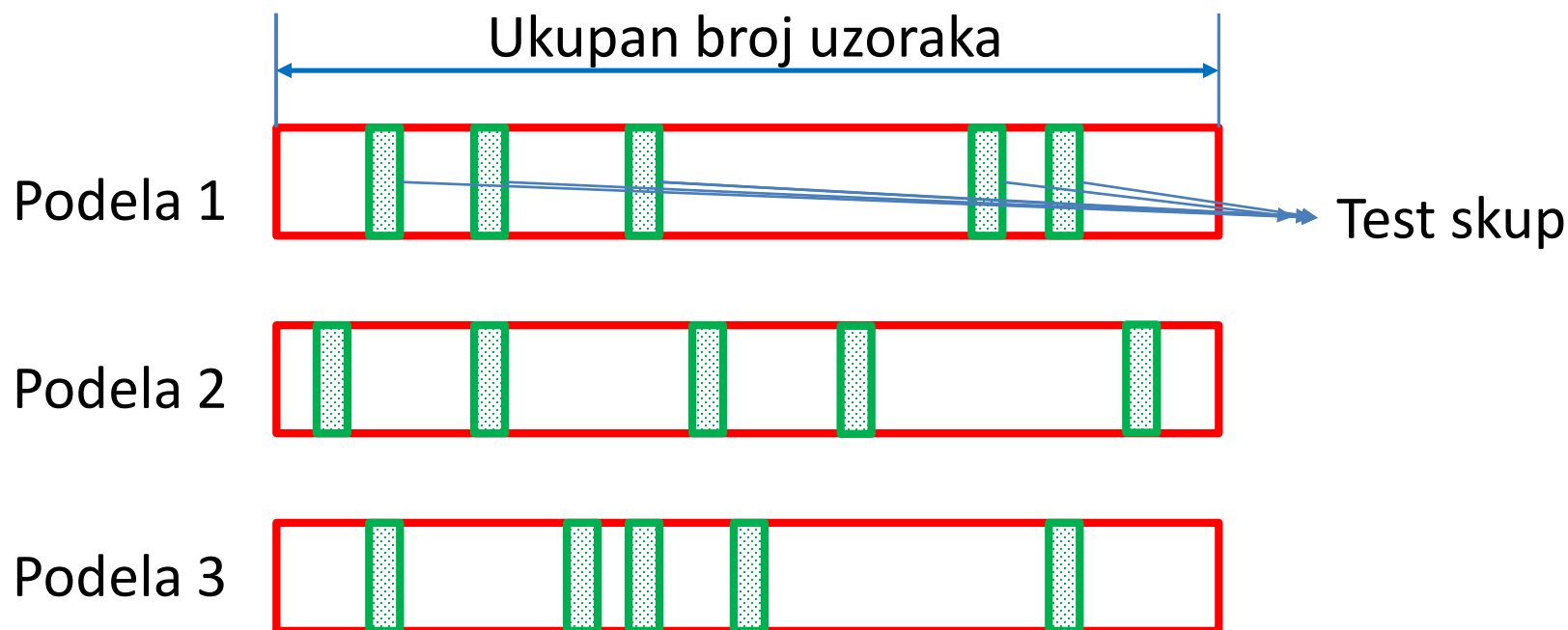
Holdout metod

- Holdout metod ima dva ozbiljna nedostatka
- U slučaju kada raspoložemo sa malo podataka, odvajanje podataka samo za testiranje predstavlja neracionalni luksuz
- Budući da se radi o jednom eksperimentu obučavanja i testiranja, holdout procena je po pravilu vrlo nepouzdana usled varijabilnosti trening i test skupa.
- Nedostaci holdout metoda se mogu prevazići korišćenjem složenijih resempling metoda uz cenu veće računske kompleksnosti.

Opšta struktura resampling metoda



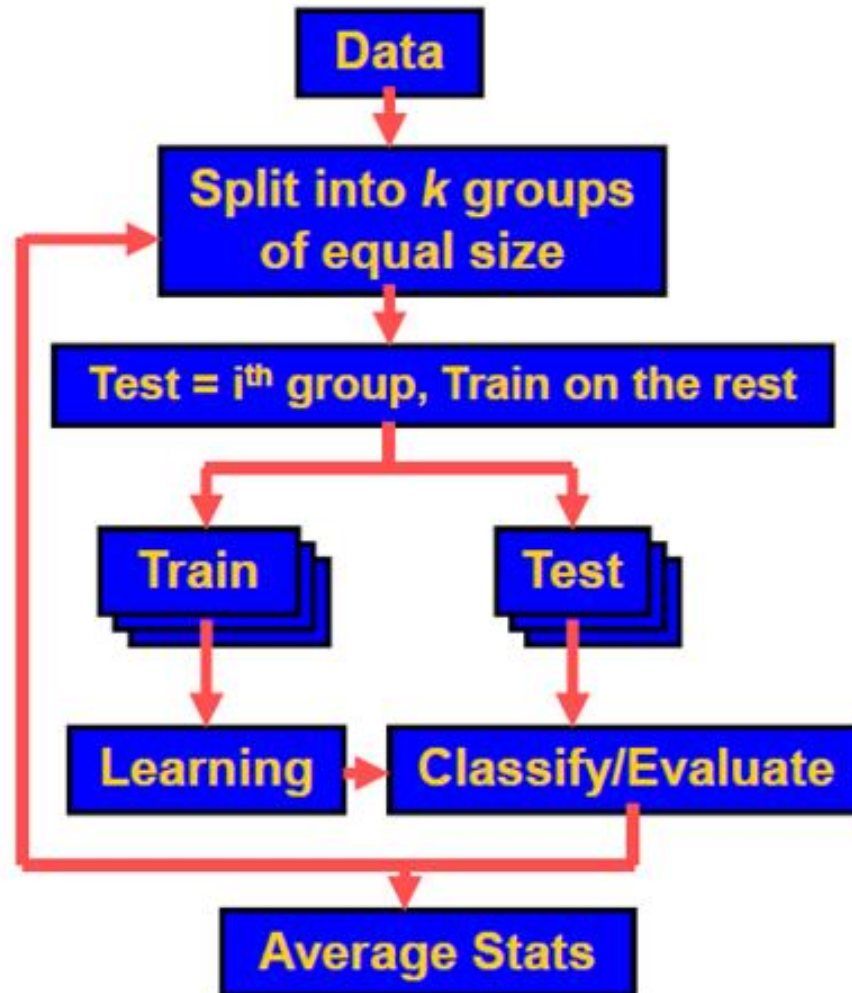
Slučajno proredjivanje (subsampling)



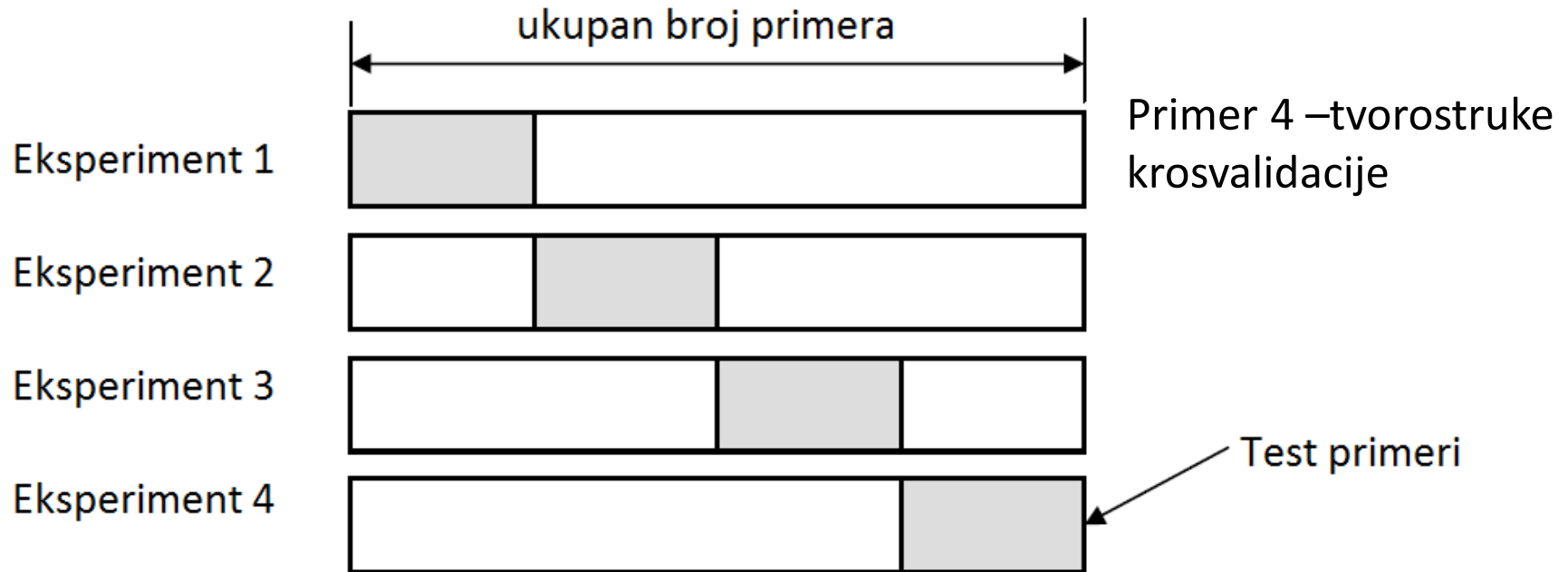
- Slučajnim proredjivanjem se od jednog obučavajućeg skupa formira K različitih podela na trening i test skup.
- Svaka podela se dobija slučajnom selekcijom fiksnog broja primera bez vraćanja.
- Finalna procena je usrednjena vrednost pojedinačnih procena

$$E = \frac{1}{K} \sum_{i=1}^K E_i$$

K - tostruka krosvalidacija



K - tostruka krosvalidacija



Kreira se podela podataka na K jednakih delova.

Za svaki eksperiment se koristi K-1 delova za trening i preostali za testiranje.

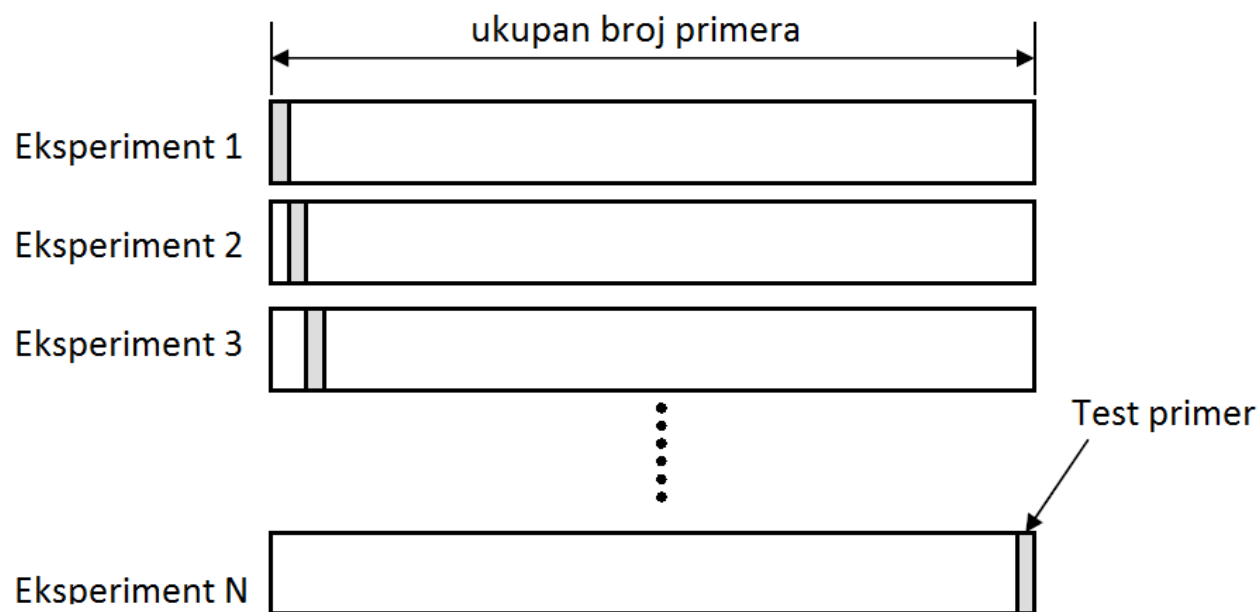
Prednost se sastoji u tome što su svi primeri na kraju upotrebljeni i za trening i za testiranje.

Finalna procena je usrednjena vrednost pojedinačnih procena

$$E = \frac{1}{K} \sum_{i=1}^K E_i$$

LOO (Leave-one-out) krosvalidacija

- LOO krosvalidacija je poseban slučaj K-tostruke krosvalidacije u kojoj je $K=N$, gde je N ukupan broj primera.



- Finalna procena je usrednjena vrednost pojedinačnih procena

$$E = \frac{1}{K} \sum_{i=1}^K E_i$$

Koliko treba da bude K

- Sa velikim K
 - ✂ Pomeraj procene greške je mali (estimator greške je veoma precizan)
 - ✂ Varijansa procene greške je vrlo velika
 - ✂ Računska kompleksnost vrlo velika (veliki broj eksperimenata)
- Sa malim K
 - ✂ Računska kompleksnost se smanjuje (manji broj eksperimenata)
 - ✂ Varijansa procene greške je mala
 - ✂ Pomeraj procene greške je vrlo velik (tzv. konzervativna procena greške, tj. veća od istinite vrednosti)
- Uobičajeni izbor u praksi je $K=10$.

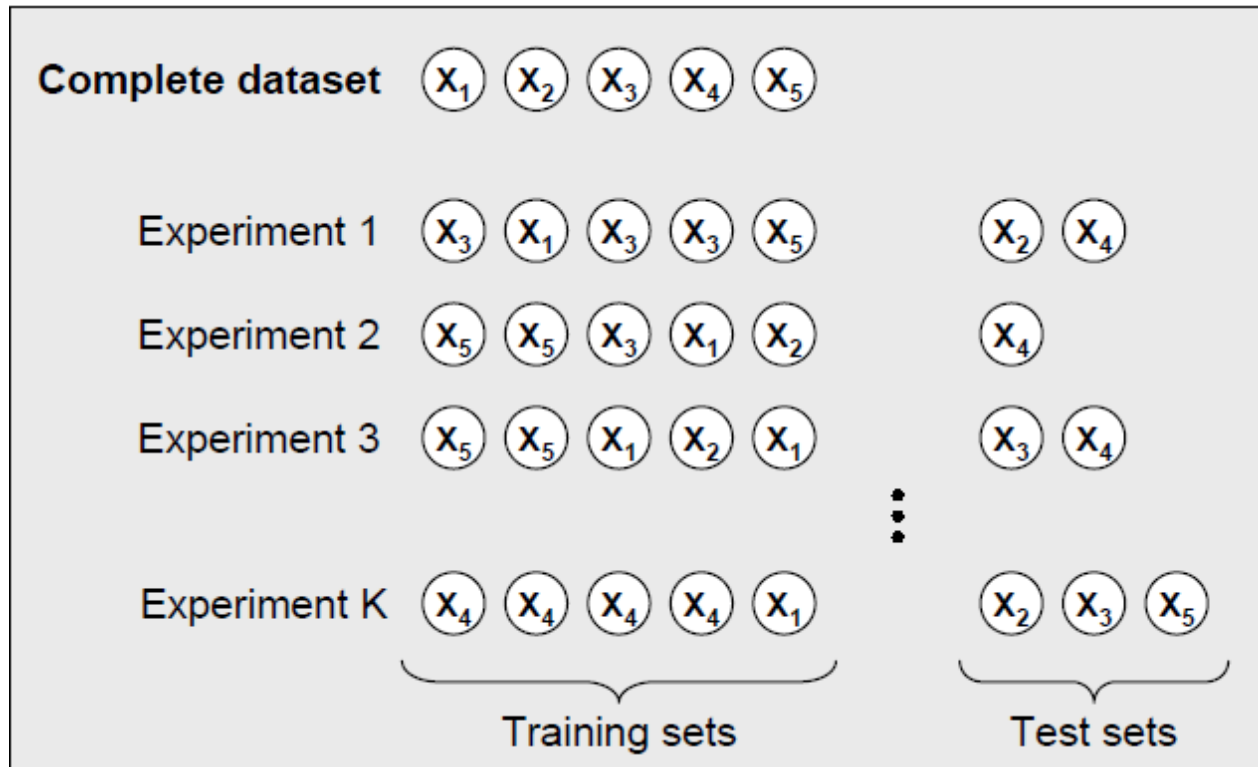
Butstrep

- Butstrep je tehnika resemplovanja sa zamenjivanjem (ponavljanjem).

Iz skupa podataka od N primera

1. Formirati trening skup slučajnim izabrom N primera sa ponavljanjem
2. Preostali primeri, koji nisu izabrani za trening skup, formiraju test skup.
3. Procena greške se dobija usrednjavanjem grešaka na test skupovima.

Butstrep



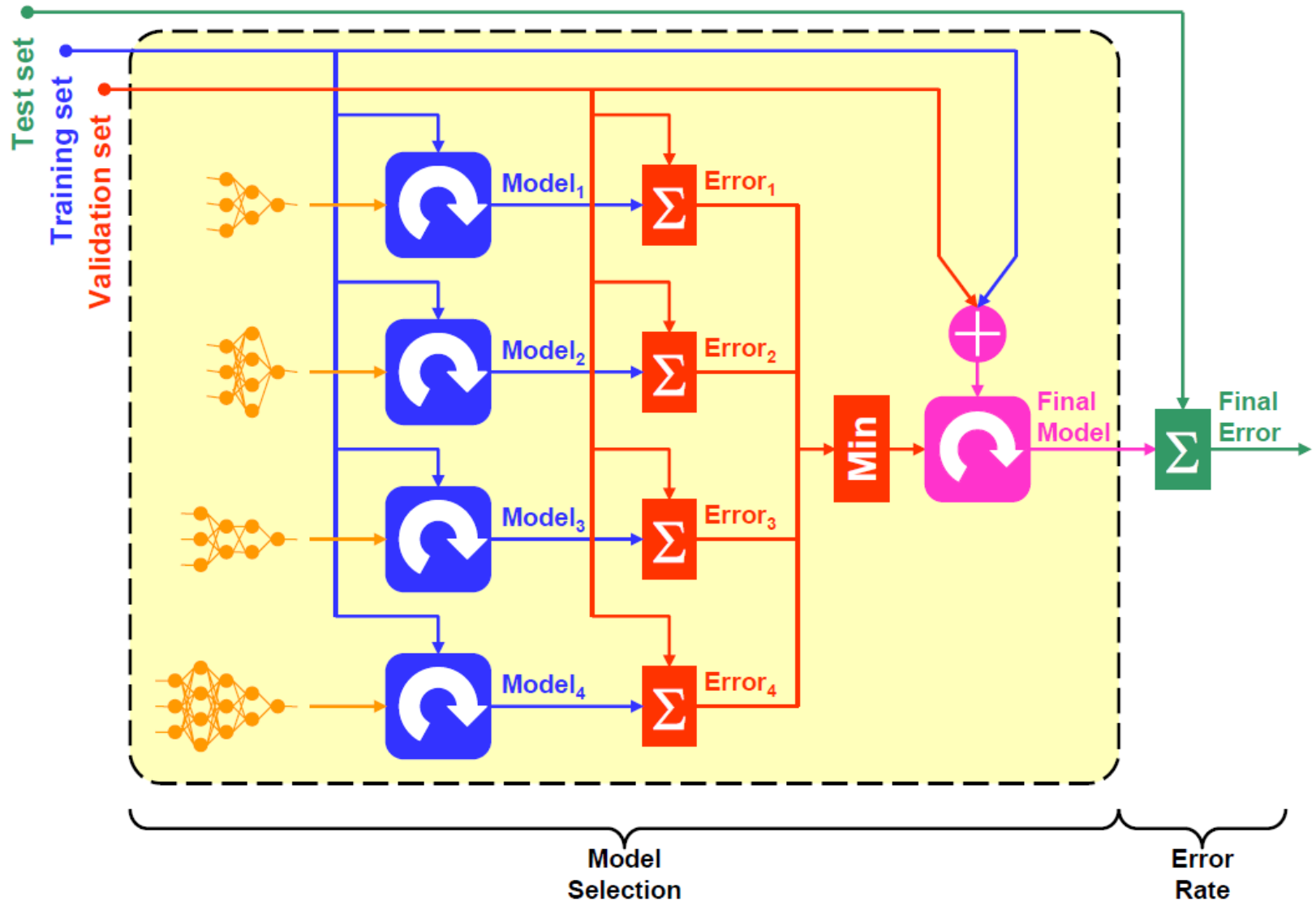
Butstrep

- U poredjenju sa osnovnim postupkom krosvalidacije, butstrep povećava varijansu. Ovo je poželjno svojstvo, budući da je realnija simulacija realnog eksperimenta na osnov koga su sakupljeni analizirani podaci.
- Resampling sa zamenjivanjem, kakav se koristi u butstrepu, ne menja apriorne verovatnoće klasa tokom procedure resemplovanja.
- Dodatna prednost butstrepa je svojstvo dobijanja precizne procene i pomeraja i varijanse greške klasifikatora.

Podela na tri skupa

- Ukoliko se istovremeno estimira i model i njegova tačnost, potrebno je dati skup podataka podeliti na tri disjunktne dela.
 - ✎ Trening skup – služi za obučavanje (procena parametara klasifikatora)
 - ✎ Validacioni skup – skup podataka koji služi za podešavanje parametara klasifikatora
 - ✎ Test skup – skup podataka koji služi samo za procenu tačnosti potpuno obučenog klasifikatora. Nakon procene tačnosti finalnog modela na test skupu, ne sme se vršiti dalje podešavanje modela.

Podela na tri skupa



Podela na tri skupa

Procedura istovremene procene modela i tačnosti:

1. Podeliti raspoložive podatke na trening, test i validacioni skup
2. Izabrati arhitekturu i parametre obučavanja
3. Obučiti model na trening skupu
4. Proceniti tačnost modela na validacionom skupu
5. Ponoviti korake 2 – 4, koristeći različite arhitekture i parametre obučavanja
6. Selektovati najtačniji model i ponovo ga obučiti na združenim podacima trening i validacionog skupa
7. Proceniti tačnost finalnog modela na test skupu

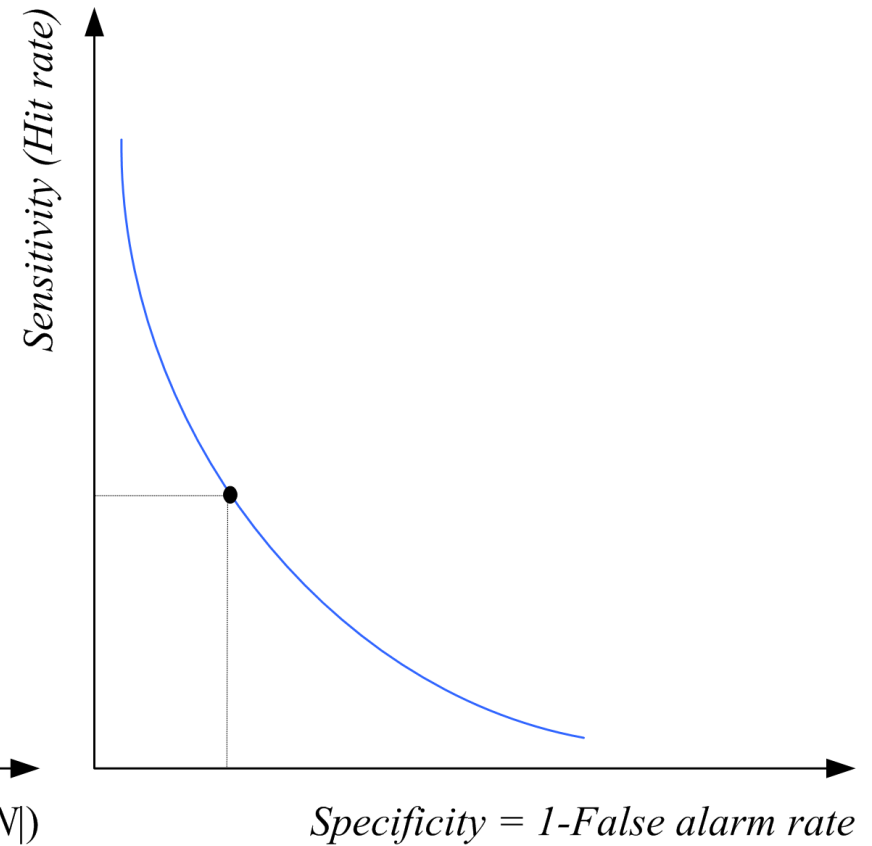
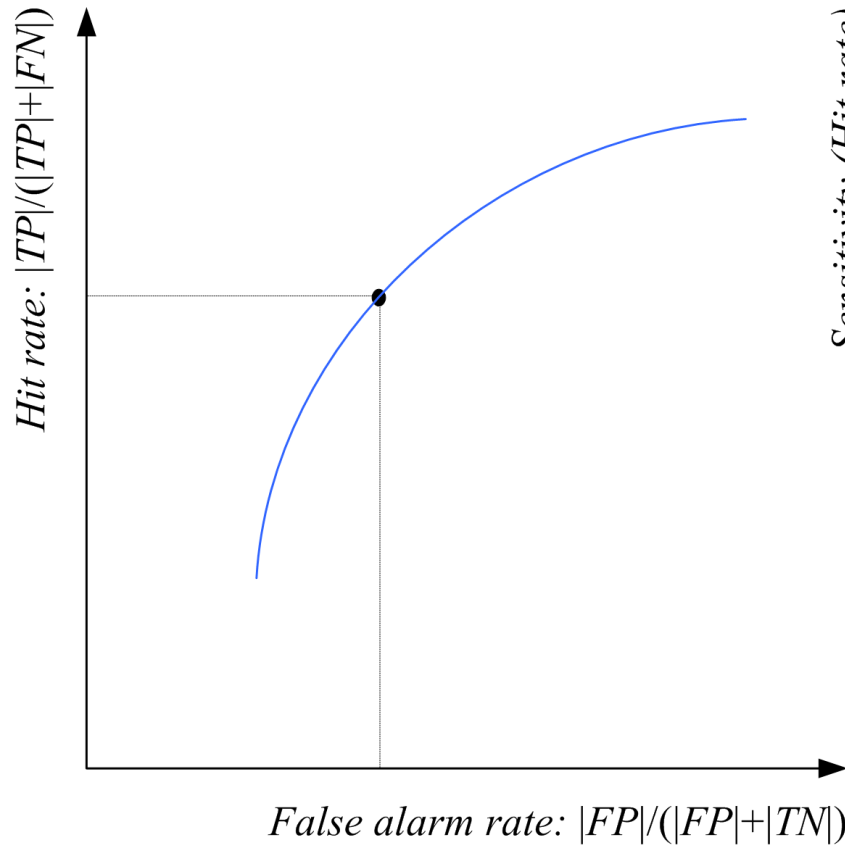
Gornja procedura podrazumeva holdout metod. Ukoliko se koristi krosvalidacija ili bootstrap, korake 3 i 4, treba ponoviti za svako K.

Mere performansi ML sistema

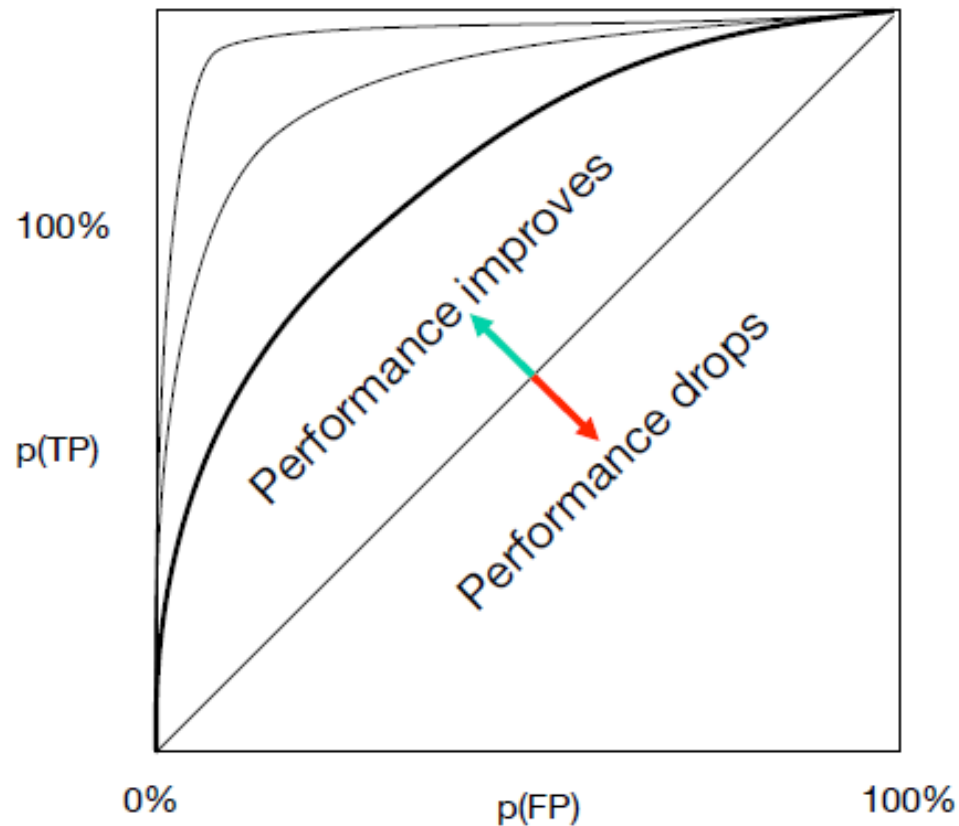
True Class	Predicted class	
	Yes	No
Yes	TP: True Positive	FN: False Negative
No	FP: False Positive	TN: True Negative

- Error rate = # of errors / # of instances = $(FN+FP) / N$
- Recall = # of found positives / # of positives
= $TP / (TP+FN)$ = sensitivity = hit rate
- Precision = # of found positives / # of found
= $TP / (TP+FP)$
- Specificity = $TN / (TN+FP)$
- False alarm rate = $FP / (FP+TN)$ = 1 - Specificity

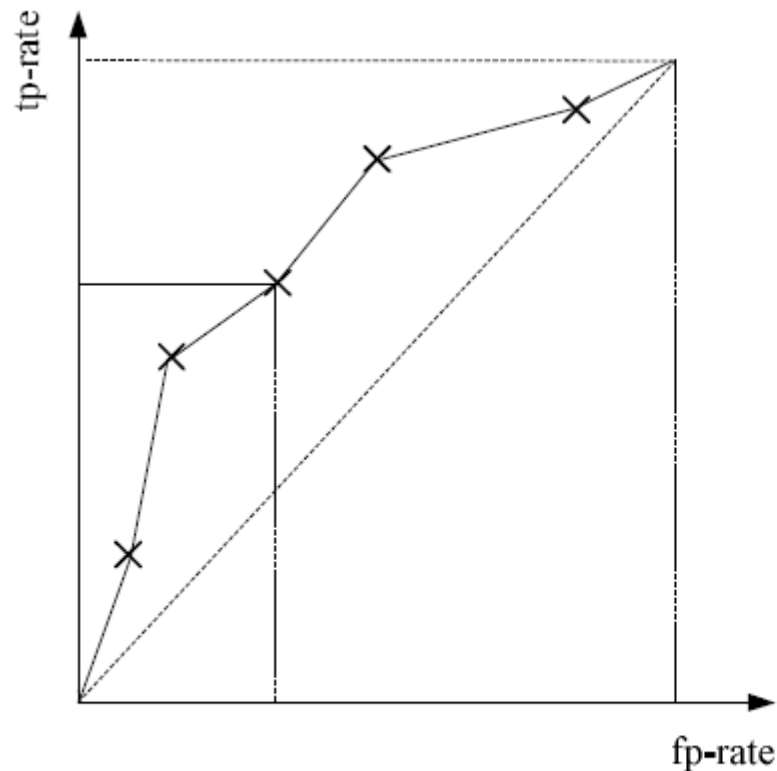
Receiver Operating Characteristic -ROC



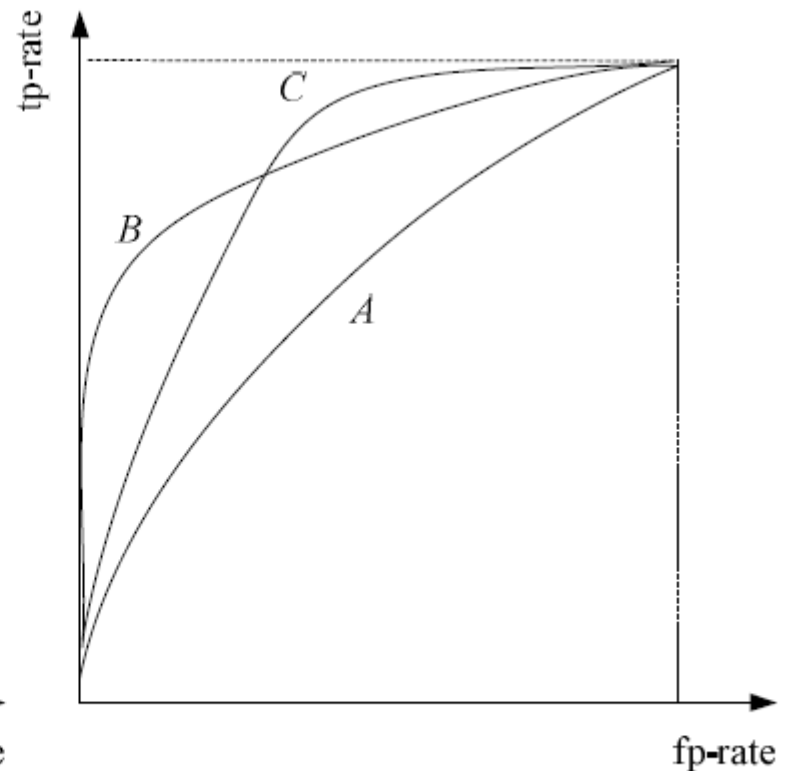
Receiver Operating Characteristic -ROC



Receiver Operating Characteristic -ROC

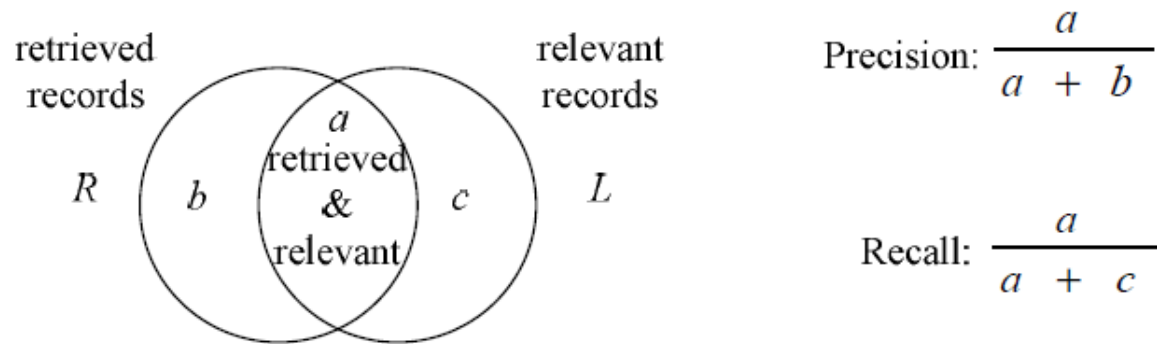


(a) Example ROC curve

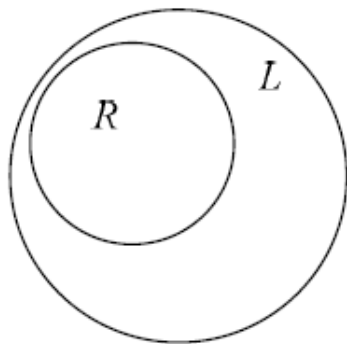


(b) Different ROC curves for different classifiers

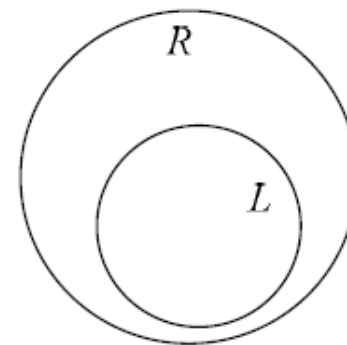
Precision and Recall



(a) Precision and recall



(b) Precision = 1



(c) Recall = 1