Non-Parametric model: Parameters keeps on adding.
↓
Doesn't mean it has not Parameters.

$$P(x) = \int_{\theta} P(x|\theta)\,\phi(\theta)\,d\theta$$

Conjugate priors:

$\boxed{\text{Gaussian}} \leftrightarrow \boxed{\text{Gaussion}}$
↓
Data comes from gaussion & mean of gaussion also comes from gaussion.

Multinomial Distribution: models the probability of counts for each of "k"-sided die rolled n-times.
→ generalization of binomial distribution.

Lecture-42 Evaluation Measures 1

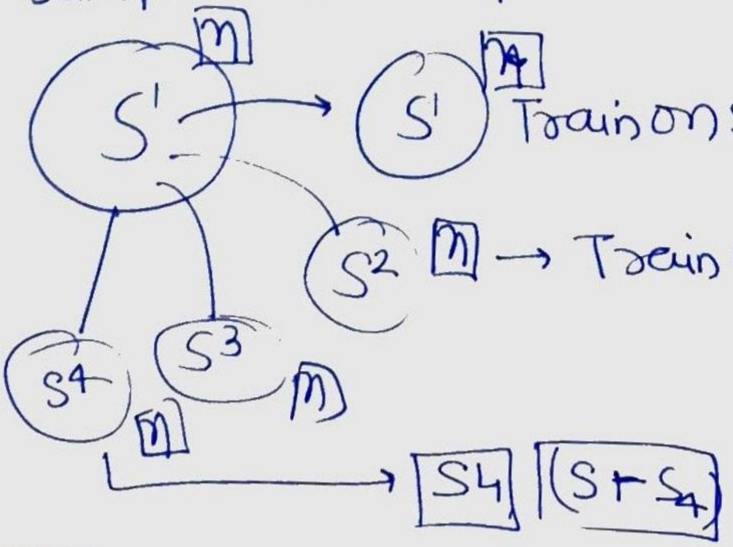
why to divide Training datasets into multiple sets

- because - Variance of model decreases, if we train it on multiple datasets.

Given 'n' independent observations $Z_1, Z_2, \ldots Z_n$, each with variance σ^2, the variance of mean $\bar{Z}$ of the observation is given by $\frac{\sigma^2}{n}$.

Bootstrapping & cross-validation

↓
sample with replacement.



Train on S1 and test on (S-S1)

→ Train on S2 and test on (S-S2)

K- FOLD CROSS VALIDATION

which one gives better estimate? bootstrap/ cross validation

No. of samples are sufficiently large: bootstrap.(67% Train)

" " " " " low : cross validation

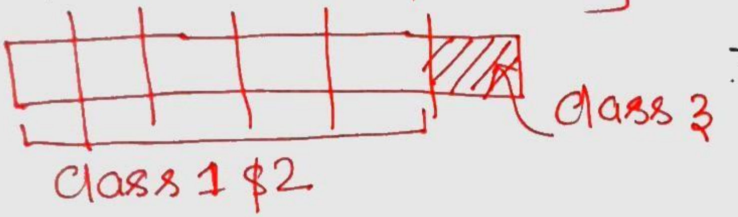
bootstrap is basically used to estimate the variance of model. like (mean of distribution.)

variation of

$\boxed{K}$, PROCESS of creating 'k' → (5 ~~to~~ or 10)

larger k ⇒ More reduction in variance

but increasing k ⇒ More biasness., ⇒ we need to choose 'K' using bias/variance reduction.



⇒ 100% error

STRATIFIED SAMPLIN: AFTER SAMPLING, Data should have same proportion in all Folds.

before CV.

IMBALANCE DATASET: Fewer No. of folds.

since Minority class population is very low.

our models needs to have good amount of minority class in Training DATA.