Process of making new data based on observed data.

Resampling.

Cross validation
⇓
Used for model assessment.

Bootstrapping.
⇓
Used to provide a measure of accuracy of a parameter estimate.

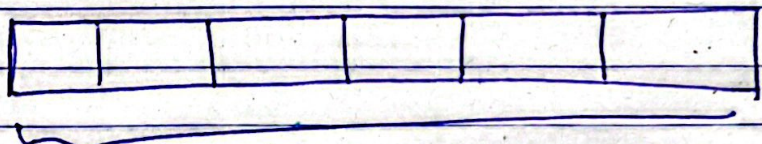
* Cross validation.

⊙ Several iterations of different splits, and
⊙ combine (typically averaging) the model accuracy across the iterations.

K-Fold cross validation.

* Divide ~~the~~/partition samples into K(near) equal sized subsamples (referred to as folds).

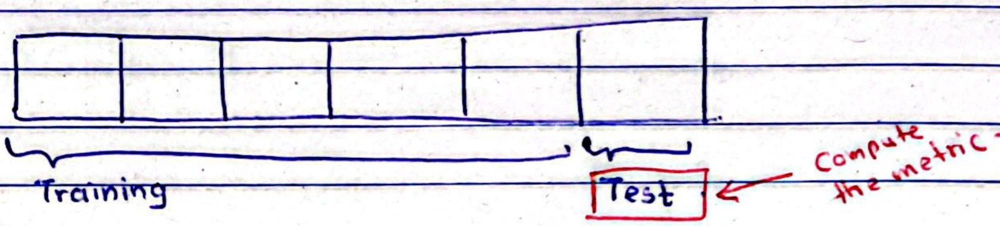
eg:-



There are 6 folds here.

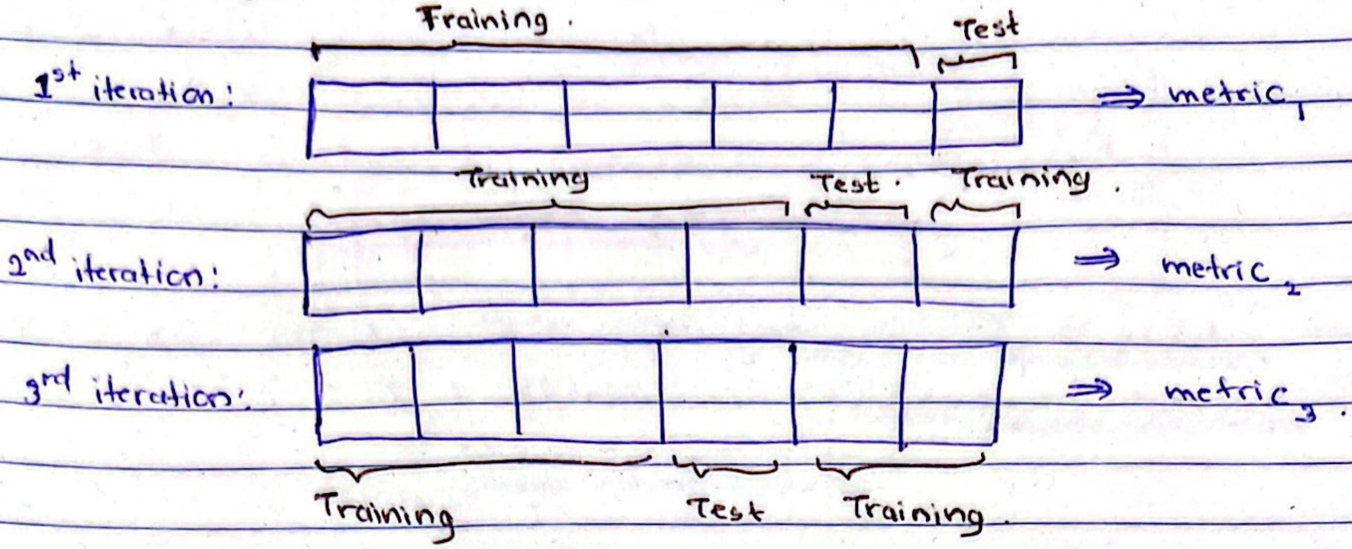
* Fit the model on (k-1) subsets, and compute a metric.
eg:- RMSE on the omitted dataset.

eg:-



* Repeat k times omitting a different subset each time.

eg:- Number of iterations = 3.



* $$\text{Cross validation accuracy} = \frac{\text{Accuracy on every fold}}{\text{Number of iteration}}$$

eg:- $$\text{cross validation accuracy} = \frac{\sum_{i=1}^{3} \text{metric}_i}{3}$$

Bootstrapping

* A bootstrap samples is created by sampling with replacement the original data with the same dimension as the original data.

Scanned with CamScanner

* Why we need Regularisation (Shrinkage) methods ?

* In subset selection methods,
  - are not guranteed to provide the best model,
  - can be slow to compute if p is large, and
  - have issues when $p > n$.

- Shrinkage methods : fit a model containing all p predictors using a technique that constrains or regularizes the coefficient estimates.

$\Longrightarrow$ This shrinks some of the coefficient estimates towards zero.

$\boxed{\text{Shrinkage methods}}$

Ridge regression. (L2 regularisation)	Lasso. (L1 regularisation).	Elastic Net. (Linear combination of L1 and L2 regularisation)
$\hat{\beta}_{ridge} = \underset{\beta \in \mathbb{R}^{p+1}}{\arg\min}\ RSS(\beta)$. subject to $\sum_{j=1}^{p} \beta_j^2 \leq s$.	$\hat{\beta}_{lasso} = \underset{\beta \in \mathbb{R}^{p+1}}{\arg\min}\ RSS(\beta)$ subject to $\sum_{j=1}^{p} \lvert\beta_j\rvert \leq s$.	
* If $s = 0$, then $\beta_1, \beta_2, \ldots, \beta_p$ are equal to zero.	* If $s=0$, then $\beta_1, \beta_2, \ldots, \beta_p$ are equal to zero.	
* If $s = \infty$, then ridge regression = OLS.	- If $s = \infty$, then lasso = OLS.	

Ridge regression

* Shrinkage is not applied to intercept term.

Form 2

$$\hat{\beta}_{ridge} = \underset{\beta \in \mathbb{R}^{p+1}}{\arg\min} \left[RSS(\beta) + \lambda \overbrace{\sum_{j=1}^{p} \beta_j^2}^{\text{Shrinkage penalty}} \right]$$

$\boxed{\lambda \geq 0}$ is called tunning parameter

⊙ If $\lambda = 0$, ridge regression = OLS.

* λ is determined typically by cross validation.

* As $\lambda \to \infty$, $\hat{\beta}_1, \dots, \hat{\beta}_p$ will approach zero.

* While it shrinks coefficients towards zero, it does not set it exactly to zero.

Lasso

Form 2

$$\hat{\beta}_{lasso} = \underset{\beta \in \mathbb{R}^{p+1}}{\arg\min} \left\{ RSS(\beta) + \lambda \sum_{j=1}^{p} |\beta_j| \right\}$$

Elastic Net.

Form 2

$$\hat{\beta}_{elastic} = \underset{\beta \in \mathbb{R}^{p+1}}{\arg\min} \left\{ RSS(\beta) + \lambda \left(\frac{1-\alpha}{2} \sum_{j=1}^{p} \beta_j^2 + \underbrace{\alpha}_{\text{Another tuning parameter.}} \sum_{j=1}^{p} |\beta_j| \right) \right\}$$

* If $\alpha = 1$, elastic net = lasso.
* If $\alpha = 0$, elastic net = ridge regression.
* You can pick λ and α via cross-validation.