

# Cross Validation

MSCI:9110 Advanced Analytics

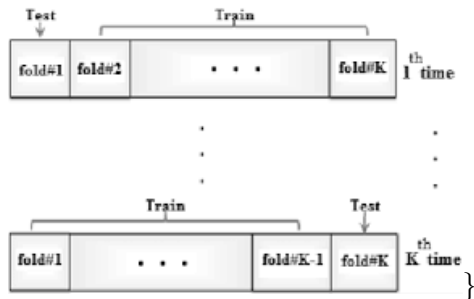
## Cross Validation

- ▶ Shortcoming of using Hold-Out Samples (like we have been doing) is
- ▶ Random Selection of a Test and Validation could result in a poor model Entirely by Chance
- ▶ To avoid this problem we have More Hold Out Samples:  $k$  rather than 1

## K-fold Cross Validation

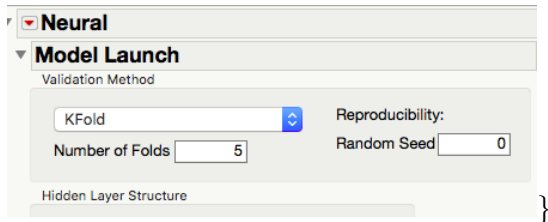
- ▶ Run  $k$  times with  $k$  separate Hold-Out sets
- ▶ During Model Building, every time a decision is made about adding more Model Complexity, entire dataset is Divided into  $k$ -subsets - the “folds”
- ▶ One of the folds is Held Out as Validation Subset while the other  $k-1$  folds are used for training
- ▶ Each fold has the opportunity to be the Validation Subset

# The K-folds



## K-fold Cross Validation with JMP Neural model

- ▶ Do Not select the Validation column
- ▶ Specify k



The screenshot shows the 'Neural' model launch dialog in JMP. The 'Model Launch' section is expanded, showing the 'Validation Method' set to 'KFold'. The 'Number of Folds' is set to 5, and the 'Random Seed' is set to 0. The 'Hidden Layer Structure' section is partially visible at the bottom.

**Neural**

**Model Launch**

Validation Method

KFold

Reproducibility:

Number of Folds 5

Random Seed 0

Hidden Layer Structure