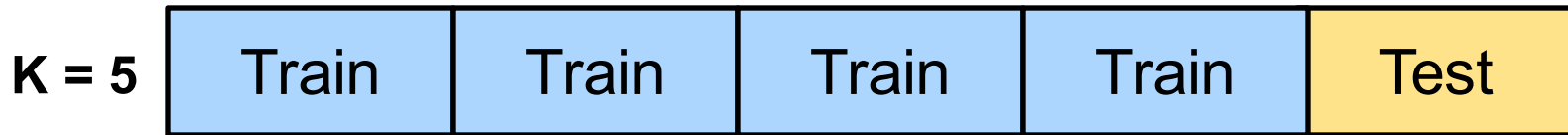


K-Fold Cross Validation (KFCV)

K-Fold Cross Validation



- It is a **widely used** approach for testing error
- The idea is to randomly **divide** the data into **K** (roughly) equal **parts**
- The model is **fit** on **K-1** folds and **tested** on the **remaining** partition
- The process is **repeated** rotating each of the **K** partitions as the holdout test fold
- The resulting **MSE** is an **average** of each of the K MSE calculations
- There is **no rule** of thumb about the most appropriate value of **K**
- A **small K** yields **small** training **samples**, so the resulting estimator coefficients are not very reliable (**K = 2** is equal to random splitting)
 - **high bias, small variance**
- A **large K** is computationally expensive and may over fit the data – **small bias, high variance**
- **K = 10** is a most popular value because it allows training the model on **90%** of the **data** each time





KOGOD SCHOOL
of
BUSINESS

