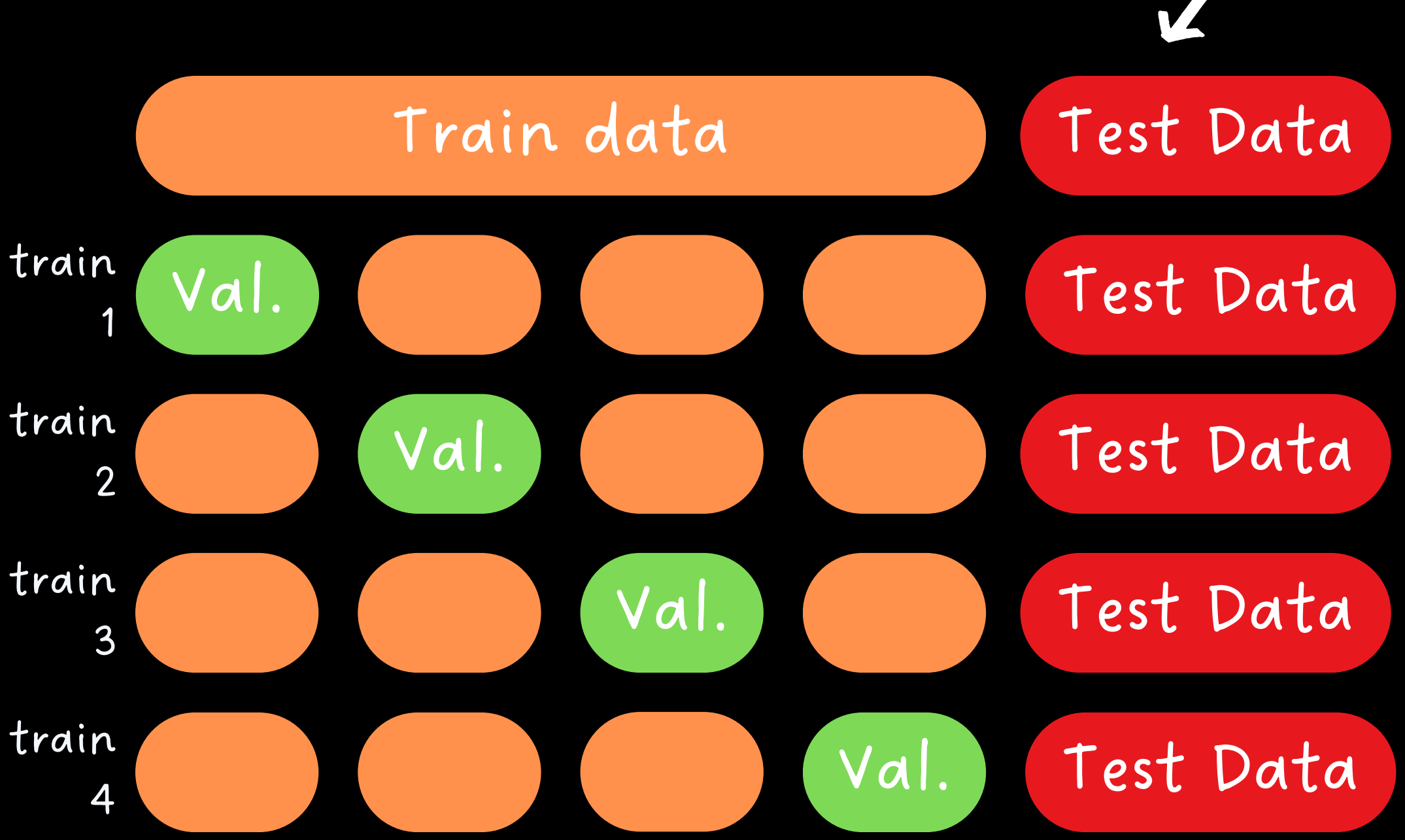
K-folds cross validation

Never use test set
except for final eval.

Here 4-folds cross validation

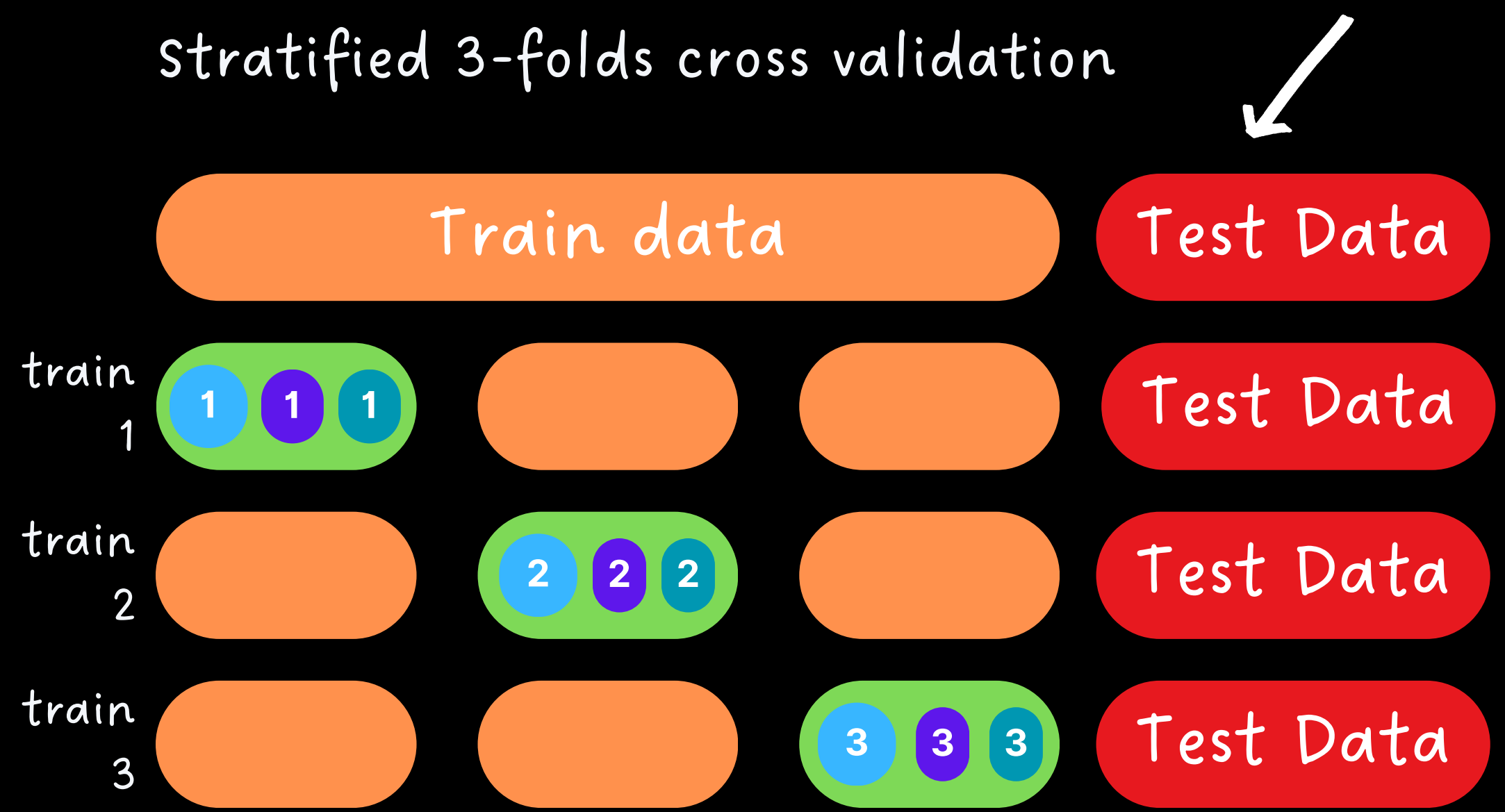


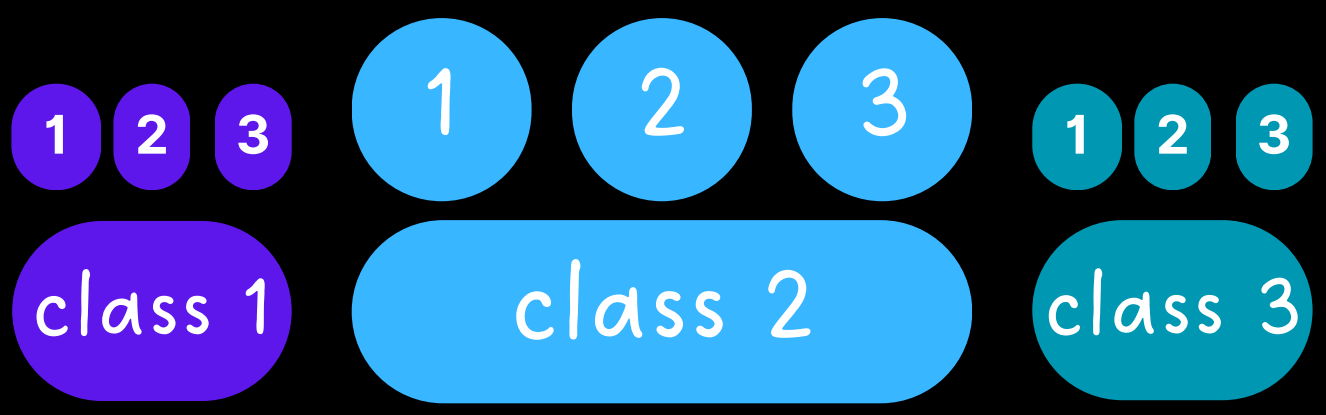
It may happen that validation set contains
only one class of the target variable for example
-> solution: stratified K-Folds in next slide.

Stratified K-folds

Never use test set
except for final eval.

Stratified 3-folds cross validation





Stratified K-Fold ensures that each fold mirrors
the class distribution of the training dataset.

Interpretation

(Non exhaustive list)

high accuracy consistently across all folds valid and training (or other performance metrics) : it indicates good generalizability!

high accuracy on training set but poor on validation set means certainly overfitting.

Low accuracy on training set and low perf. on validation set means certainly undefitting.

Limitations

Deep learning models does not require cross validation for 2 reasons:

1. Deep learning deals with larger datasets where the training set might already be representative of the population, diminishing the necessity for cross-validation.
2. Cross validation requires multiple trainings for the same model. For deep learning it may be too resource intensive and too long to train several time the same model.

Thank you for Reading

I hope this helps.