



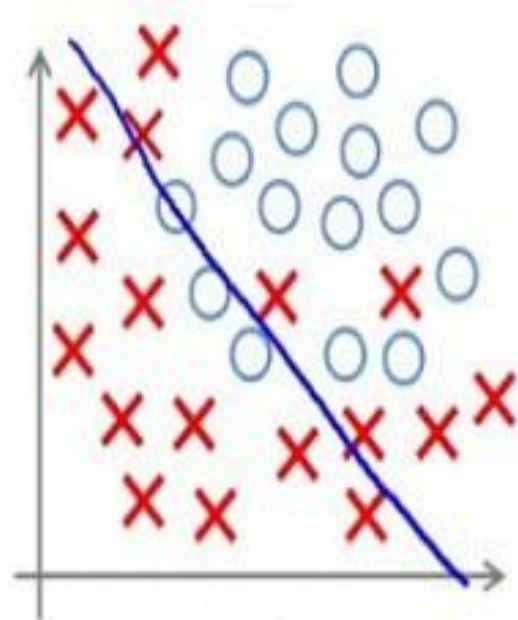
StepUp
Analytics

K Fold Cross Validation



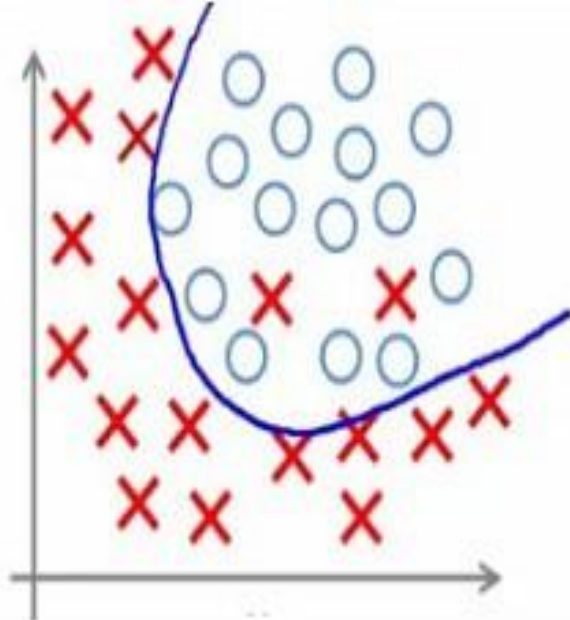
Problema de overfitting

- Modelo que modela os dados de treinamento “bem de mais”.
- Ruídos nos dados de teste podem impactar negativamente a avaliação de novos dados.
- Afeta modelos não-parametrizados mais do que problemas não parametrizados.

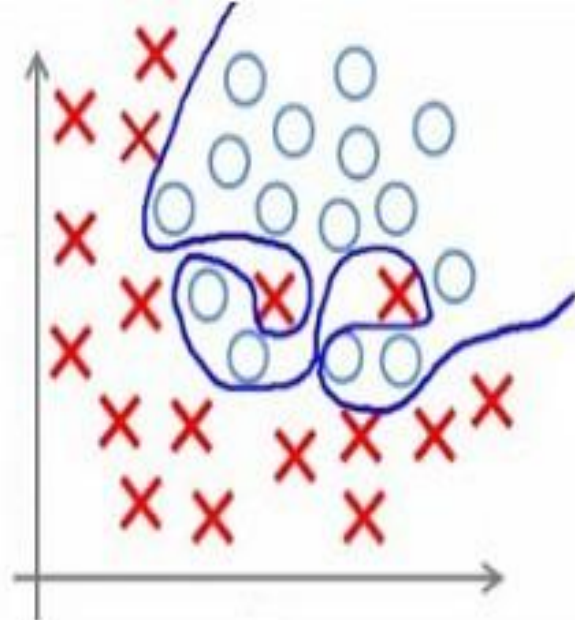


Under-fitting

(too simple to
explain the
variance)



Appropriate-fitting



Over-fitting

(forcefitting -- too
good to be true)

Cross Validation

The diagram illustrates the cross-validation process. It features a large vertical rectangle on the left labeled 'Dataset'. To its right, five smaller rectangles are stacked vertically, labeled 'SET A', 'SET B', 'SET C', 'SET D', and 'SET E'. Arrows point from the 'Dataset' box to each of these five sets. To the right of these sets, there are two more vertical rectangles: 'Training Set' and 'TEST SET'. Arrows point from each of the five sets to the 'Training Set' box. A single arrow points from the 'TEST SET' box to the 'Training Set' box.

- Particionar os dados em conjuntos, onde um conjunto é utilizado para treino e o outro para teste e avaliação do desempenho do modelo.

Dataset

- Ajuda a detectar overfitting.

- Existem vários métodos de aplicação de cross validation, o mais comum é o k-fold.

K-FOLD

- Dividir o dataset em k subconjuntos (valor de k é definido previamente).
- São feitas k interações, a cada uma um conjunto formado por $k-1$ conjuntos é usados para treino e o conjunto restante é usado para teste.
- Todos os k conjuntos são testados no final.

Iteration 1



Iteration 2



Iteration 3



Iteration 4

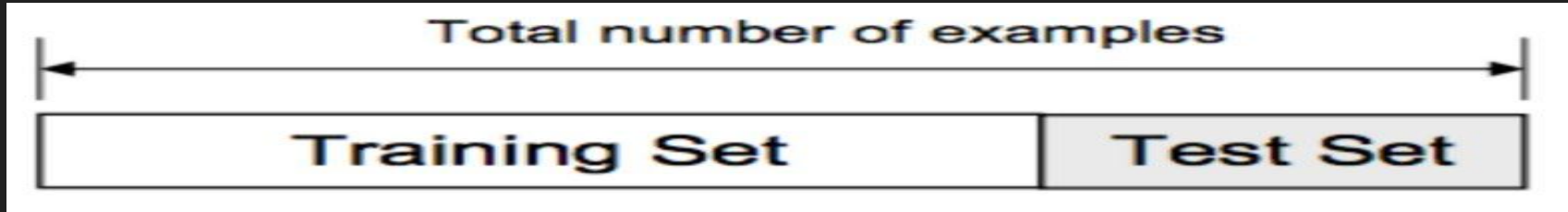


Iteration 5

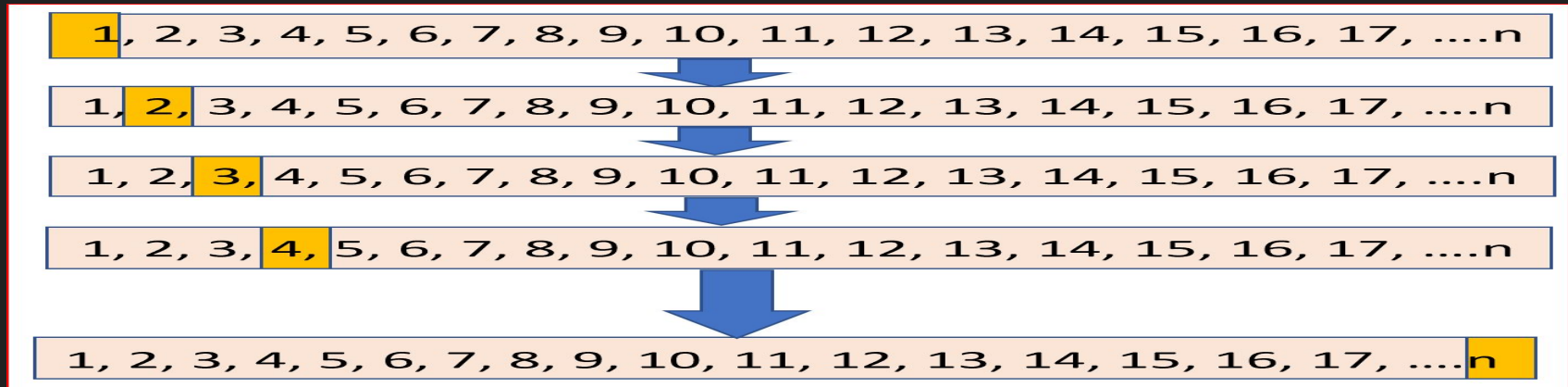


Variações do k-fold

- Train/test split.

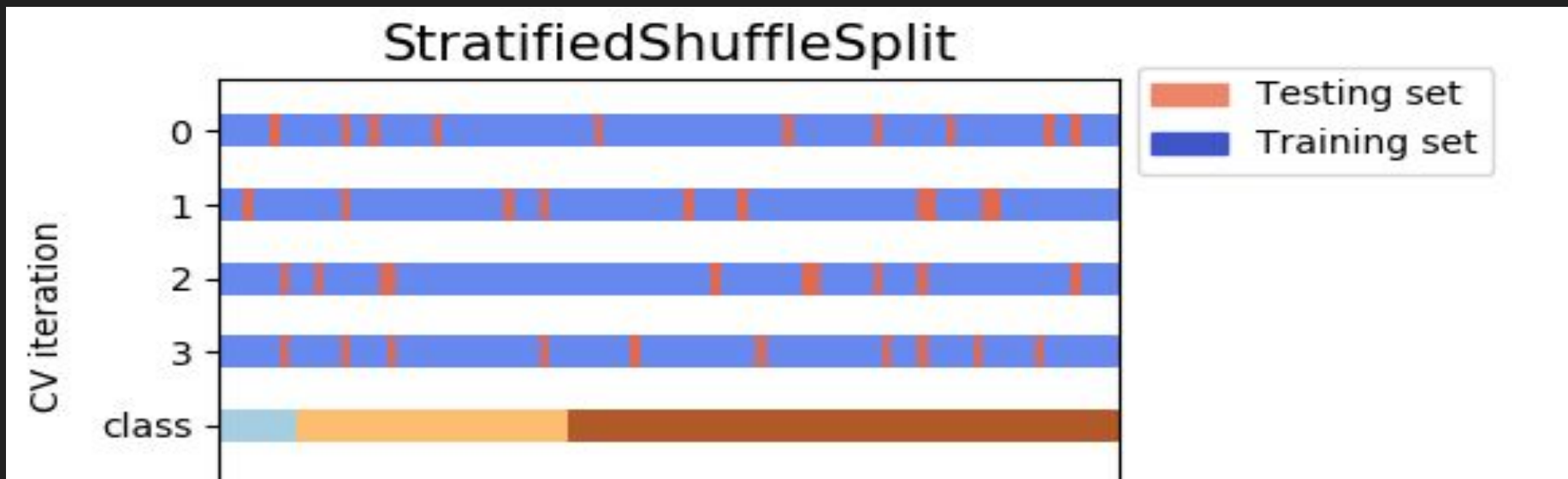


- LOOCV (leave-one-out-cross-validation).



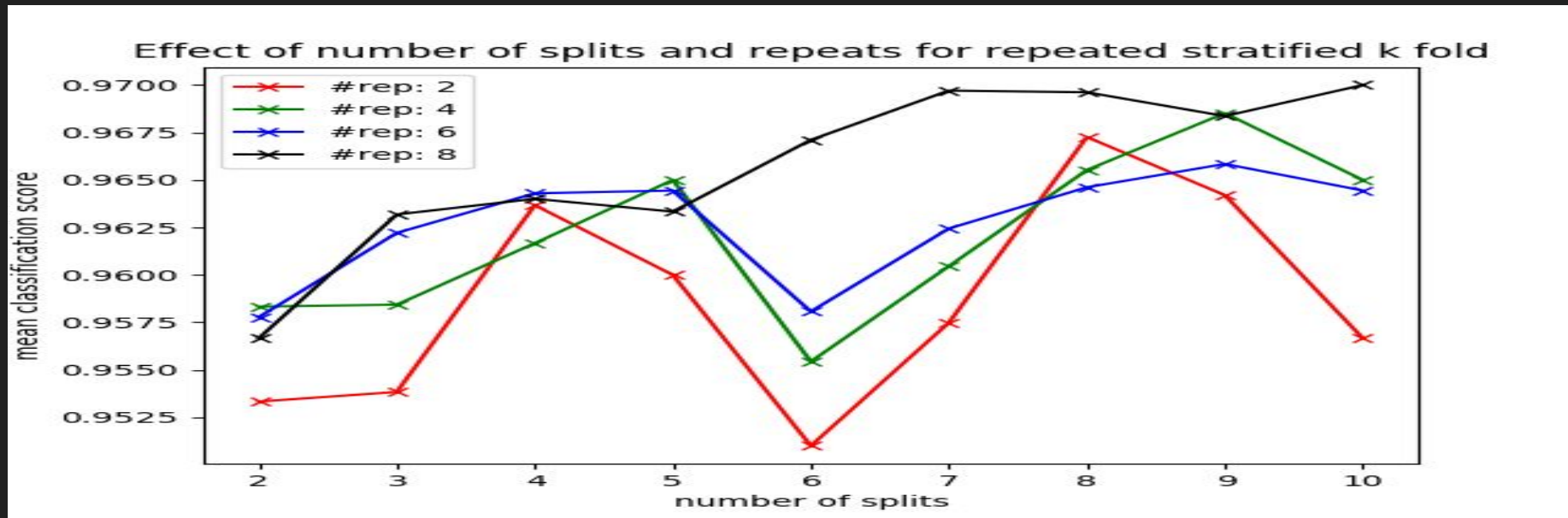
Variações do k-fold

- Stratified.



Variações do k-fold

- Repeated.



Bibliografia

- <https://machinelearningmastery.com/k-fold-cross-validation/>
- <https://medium.com/@edubrazrabello/cross-validation-avaliando-seu-modelo-de-machine-learning-1fb70df15b78>
- <https://machinelearningmastery.com/overfitting-and-underfitting-with-machine-learning-algorithms/>
- <https://machinelearningmastery.com/parametric-and-nonparametric-machine-learning-algorithms/>