Data Analytics for Personalized Genomics and Precision Medicine Data & Python

Lecturer: Yu LI (李煜) from Department of Computer Science and Engineering (CSE)

Lecture 9 – Clustering and Classification Performance Evaluation
Monday, October 3rd, 2022

Outline of the Lecture:

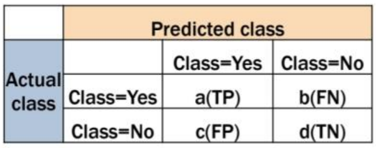
- Performance Evaluation
- Cross-Validation
- Multi-class Classification
- Clustering Evaluation

1. Performance Evaluation

- Performance Evaluation:
  - Definition: Quantify the performance to summarize the performance of the different models.
  - Purpose:

    - To characterize the performance of a model and pinpoint the strong and weak points of it.

    - To determine which method is the most suitable among the different classification methods and assist in method selection.
  - Method: Confusion Matrix
- Confusion Matrix



		Predicted class	
		Class=Yes	Class=No
Actual class	Class=Yes	a(TP)	b(FN)
	Class=No	c(FP)	d(TN)

Confusion Matrix is a table used to determine the performance of classification. It is a 2×2 matrix where each box represents True Positive (a), False Negative (b), False Positive (c), True Negative (d) respectively.

- $Accuracy = \frac{a+d}{a+b+c+d}$
- $Precision = \frac{a}{a+c}$

  Among predicted positive samples, number of correct prediction (actual positive)

- $Recall = \frac{a}{a+b}$

  Among the actual positive samples, number of correct prediction (positive)

- $F1\ score = \frac{2 * precision * recall}{precision+recall}$

  Weighted average of precision and recall

- If we have a bad classifier (responses in only Yes), it will be reflected in accuracy test.

  If we have a bad classifier (responses in only Yes) & imbalanced classes, accuracy may be misleading.

  Precision, Recall, F1 score also have imbalanced data induced problem.

- If data is imbalance, check balanced accuracy or look at the confusion matrix directly.

  $$Balanced\ accuracy = 0.5 * \left(\frac{TP}{TP + FN} + \frac{TN}{TN + FP}\right) = 0.5$$

- Because the values are not absolute, context is important in choosing the evaluation method. For example, for rare cancer pre-screening, it is best to choose a method that runs over all potential positives so that I don't miss slightest chance of cancer.

2. Cross-Validation

- Cross-fold validation:

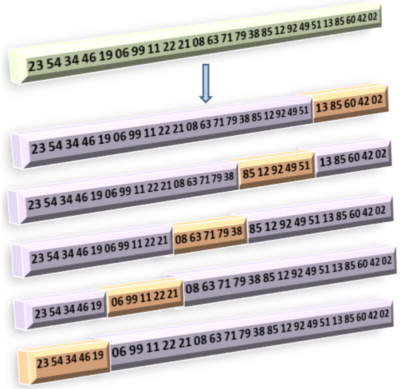
  - Purpose: To choose distance metrics and K in KNN that will result in good prediction accuracy

  - Problem: There aren't exact label for testing data

  - Solution: Use part of training data as testing data and calculate the performance. Divide the data into several different parts repetitively to calculate the average.

- N-fold cross validation: Dividing data into several fold and averaging the validation set accuracies

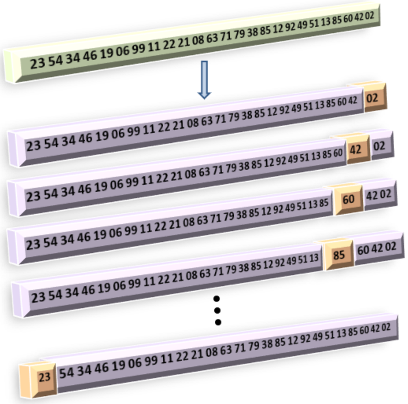
  - 5-fold cross validation:



Set one-fold among the training data as the test data and calculate the performance accordingly.

Repeat this and calculate the average.

- Leave-one-out cross-validation: special case of N-fold cross-validation where number of folds equals number of instances.



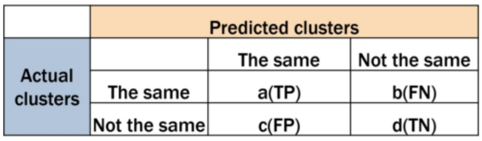
3. Multi-class Classification

- Multi-class classification:
- Multi-class evaluation: aggregating multiple values (ex. Accuracy, Precision, Recall, F1 score) into one value
  - $Macro-average = \frac{Accuracy\ 1 + Accuracy\ 2 + \cdots + Accuracy\ n}{\#\ of\ class}$

    (Low performance of small classes will show up.)
  - $weighted-average = \frac{Accuracy\ 1 * weight\ 1 + Accuracy\ 2 * weight\ 2 + \cdots + Accuracy\ n * weight\ n}{total\ weight}$

4. Clustering Evaluation

- Clustering evaluation: Checking if they are clustered (two similar cells are in the same cluster) correctly. They do not need to be classified correctly.



		Predicted clusters	
Actual clusters		The same	Not the same
	The same	a(TP)	b(FN)
	Not the same	c(FP)	d(TN)

- Rand Index R: $R = \frac{a+d}{Number\ of\ pair\ combination}$
- Number of Pairs = $Number\ of\ Pairs = \binom{n}{2} = \frac{n*(n-1)}{2}$
- Check if a real pair of cells is the same cluster. Then check the prediction about the same pair. If these two results are identical (same-same, different-different), it means that it was clustered correctly.