Lecture 10

<Clustering and Classification Performance Evaluation>

Kiwoong Nam **1155113612**

1. Cross-validation

Standard procedure of K-Nearest Neighbor (KNN)

>> Normalization

>> Compute distances

>> Identify the most similar data

>> Take their class out and find the mode class

(After the distance metric and K is chosen)

**The question is:

How do we get the best K, when all we have in hand is just the data itself, and *we do not have the label for the testing data?*

An algorithm that can be used, in this case, to determine the best K

: CROSS-VALIDATION / ROTATION ESTIMATION

(: a technique for assessing how the results of a machine learning analysis will generalize to an independent data (test data) set.)

(: it is to measure the performance of models.

)

- ➔ Use part of training data (each part one by one), pretending that it is the testing data (partitioning a set of data into complementary subsets)
- ➔ Perform standard algorithms on each subset
- ➔ Evaluation metric for each subset
- ➔ Then calculate the AVERAGE over all parts

Lecture 10

<Clustering and Classification Performance Evaluation>

Kiwoong Nam **1155113612**

*****In cross-validation, only the training data <u>with label</u> are being considered.**

Then, for final prediction, once K is determined, a classifier is applied on to the testing data

<u>Example of utilizing cross-validation</u>

<u>- If K=1:</u>

Person	Height	Weight	Gender
P1	0.625	0.875	M
P2	0	0	F
P3	0.25	0.375	M
P4	1	1	M
P5	0.4583	0.6667	??

	P1	P2	P3	P4
P1	0	0.875	0.5	0.375
P2	0.875	0	0.375	1
P3	0.5	0.375	0	0.75
p4	0.375	1	0.75	0

P1 is the closest with P4 (0.375) (as highlighted in yellow in the image), and according to the data given, P4 is M. (→ P1 predicted as M)

Same steps are taken for P2, P3 and P4.

Below image shows the results:

Lecture 10

<Clustering and Classification Performance Evaluation>

Kiwoong Nam **1155113612**

P1: P4—M
P2: P3—M
P3: P2—F
P4: P1—M

Comparing the prediction results and the data, we can derive that out of 4, two results are correct and the other two are incorrect.

➔ Accuracy = (2 correct / 4 samples) = 1/2 = 0.5

- If K=3:

Here, we take the mode class from the other samples except the one being predicted for :

Person	Height	Weight	Gender
P1	0.625	0.875	M
P2	0	0	F
P3	0.25	0.375	M
P4	1	1	M
P5	0.4583	0.6667	??

Prediction for P1:

Mode class for the other three samples (P2, P3 & P4) → M

Therefore, prediction for P1 → M

Same steps are taken afterwards and below, the image shows the prediction results:

Lecture 10

<Clustering and Classification Performance Evaluation>

Kiwoong Nam **1155113612**

P1: M
P2: M
P3: M
P4: M

Again, comparing the results with the data value, three are correct out of four.

Accuracy = 3/4 = 0.75

Since, by comparing the cross-validation accuracy, <u>the accuracy is higher when K=3, than when K=1.</u>

Therefore, K=3 is chosen.

<u>1-1: n-Fold Cross-Validation</u>

<u>n: the number of partitions</u>

: Train multiple times, leaving out a disjoint subset of data each time for validation. Then, <u>take the average</u> of validation set accuracies.

Process:

- Randomly partition data into n disjoint subsets

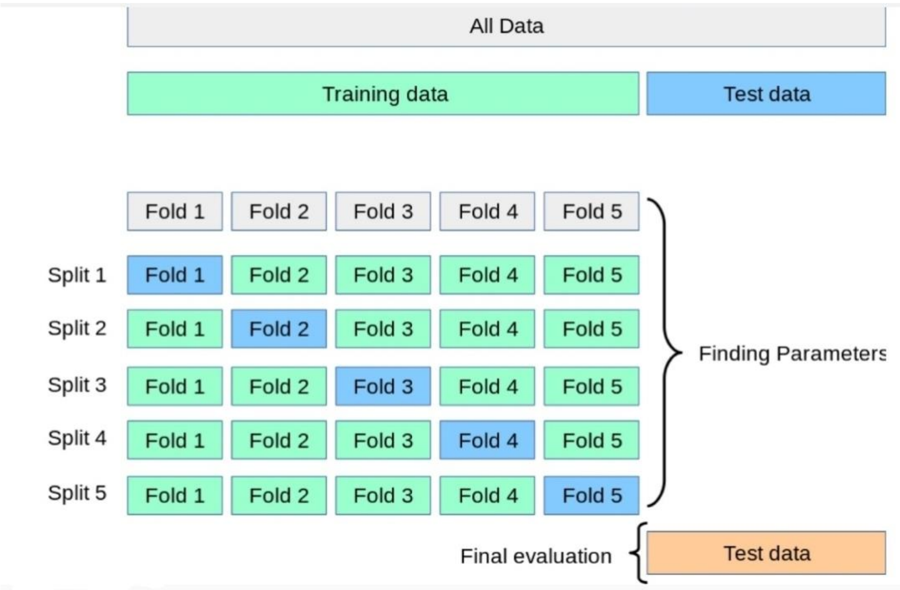
Lecture 10

<Clustering and Classification Performance Evaluation>

Kiwoong Nam **1155113612**

- For I =1 to n,
    - Validation Data = i-th subset
    - h: classifier trained on all data except for Validation Data
    - Accuracy (i) = accuracy of h on Validation Data
- Final Accuracy = mean of the n recorded accuracies

Example : 5-fold cross-validation



There are 10 data points, and 5-folds (5 different subsets are made, grouped randomly).

Let's say, of the samples ranging from P1 to P10,

P1-2, P3-4, P5-6, P7-8, P9-10 → 5 subsets/folds

Then, for example, if P1-2 is "Split 1" (in the image), consider the Fold 1 as a test data, and the other folds as training data, predict the accuracy for Split 1.

Repeat the steps for Split 2 and all the way to Split 5.

We now have five accuracy predictions.

>> Final Accuracy = the average of the five accuracies.

*** We use all the training data to train the model and then, evaluate on test data.

1-2: Leave-one-out cross-validation

This is a special case where *the number of folds is equivalent to the total number of data points.*

- ➔ In each fold, there is only one data point for evaluation.
- ➔ Other than that, everything is the same with n-fold cross-validation. (although it takes more time than n-fold cross-validation.)

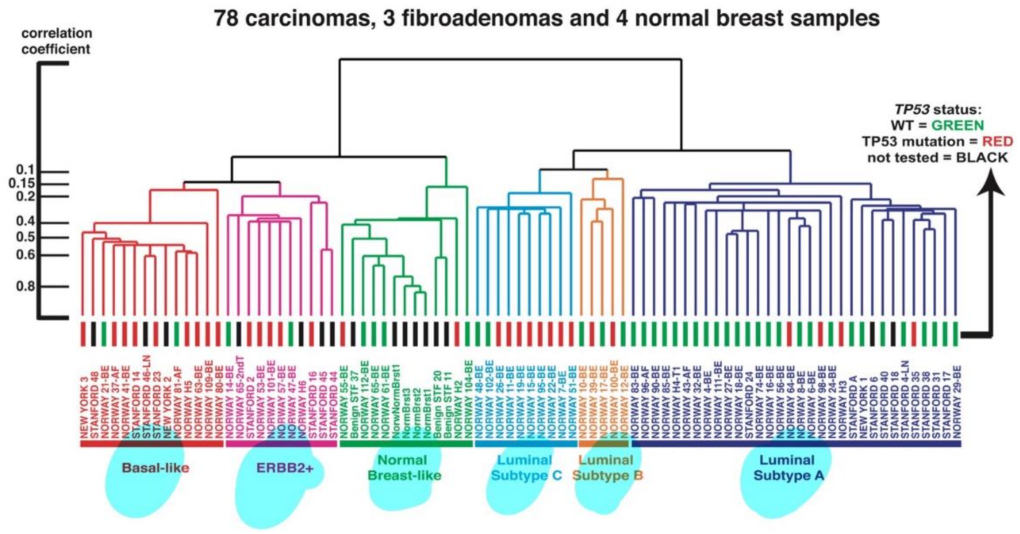
Lecture 10

<Clustering and Classification Performance Evaluation>

Kiwoong Nam **1155113612**

2. Multi-class Classification

(Intuitive idea : generate a lot of labels)



The example image shows that there are more than 2 classes.

Instead, there are 6 classes (multi-class).

>>> For KNN,

When we make predictions and take the mode class,

we can directly predict from different classes.

(Probably, in the training data, we already have 6 different labels.)

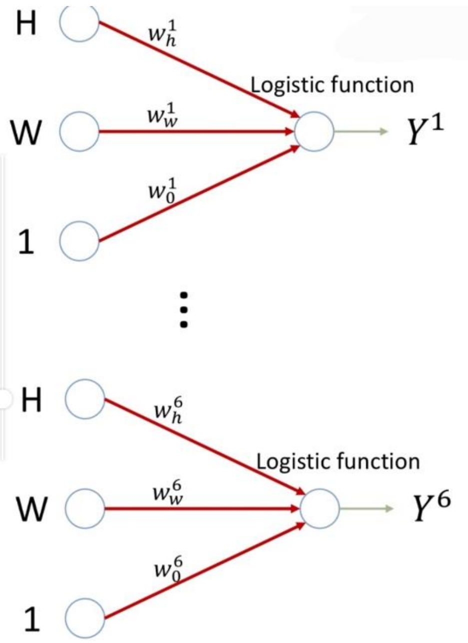
Lecture 10

<Clustering and Classification Performance Evaluation>

Kiwoong Nam **1155113612**

>>> For logistic regression,

We just have a threshold (either 0 or 1) and therefore, the output from logistic regression model would be binary.



HOW CAN WE CONSTRUCT A MULTI-CLASS CLASSIFIER?

- ➔ Instead of just one logistic regression model, build a number of logistic regression models, one for each class.
- ➔ When predicting, we assign class with the highest value.
- ➔ When training, we train 3*8 = 18 parameters.

Multi-class evaluation still uses accuracy, precision, recall, F1 score etc.

→ Consider each class as a binary classification problem.

→ Take the average of each class.

→ Use the average to assess the performance of multi-class classification models.

Lecture 10

<Clustering and Classification Performance Evaluation>

Kiwoong Nam **1155113612**

HOW DO WE AGGREGATE MULTIPLE VALUES INTO ONE VALUE?

Class	Accuracy	Cells
1	0.9	150
2	0.95	50
3	0.85	100
4	0.8	40
5	0.7	20
6	0.2	10

$$Macro - average = \frac{0.9 + 0.95 + \cdots + 0.7 + 0.2}{6} = 0.73$$

$$Micro - average = \frac{0.9 * 150 + \cdots + 0.2 * 10}{150 + \cdots + 10} = 0.85$$

Macro – Average (Unweighted mean)

: Considers each class equally.

: **Shows the performance for each class equally.**

So if the performance is not so good with small classes, then the overall performance is not so good.

Micro – Average (Weighted mean)

: **Multiplies the performance with the number of the data points within the class**

: Good performance on large class → Better performance

Lecture 10

<Clustering and Classification Performance Evaluation>

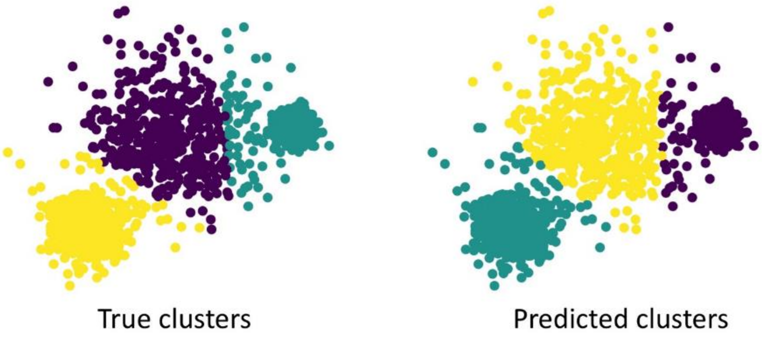
Kiwoong Nam **1155113612**

3. Clustering Evaluation

Clustering is DIFFERENT from classification.

Classification – correct, only if we predict the label correctly.

Clustering – correct, as long as the label is the same for some data points (even if the label is not predicted correctly)



The above image displays a **poor prediction in classification.**

However it displays a **great prediction in terms of clustering.**

In clustering, it is correct as long as the **two data points are in the same cluster** and **a pair of datapoints should be evaluated.**

Lecture 10

<Clustering and Classification Performance Evaluation>

Kiwoong Nam **1155113612**

TP: True Positive, FP: False Positive, TN: True Negative, FN: False Negative

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

For all the pairs in the dataset:

a: the number of pairs are in the same cluster in the True clusters and also assigned to one cluster in the Predicted clusters
b: the number of pairs are in the same cluster in the True clusters and also assigned to different clusters in the Predicted clusters
c: the number of pairs are in different clusters in the True clusters and also assigned to one cluster in the Predicted clusters
d: the number of pairs are in different clusters in the True clusters and also assigned to different clusters in the Predicted clusters

Evaluation metric used for clustering (similar to accuracy in classification)

: RAND INDEX (R)

Rand index, R:

$$R = \frac{a+d}{a+b+c+d} = \frac{a+d}{Number\ of\ all\ the\ pair\ combinations}$$

$$Pairs = \binom{n}{2} = \frac{n*(n-1)}{2}$$

n: Total number of points

Lecture 10

<Clustering and Classification Performance Evaluation>

Kiwoong Nam **1155113612**

Pairs : show how many pairs that can be there if there are n datapoints.

a + d : the number of pairs that are predicted correctly.