# Multicategory Classification

1- Given the following data, and using
   a) One-vs-all approach
   b) Pairwise approach

| Measurement | Category |
|---|---|
| 1 | D |
| 1.3 | D |
| 4.9 | A |
| 2.1 | C |
| 3.3 | B |
| 4.8 | A |
| 1.9 | C |
| 5.1 | A |

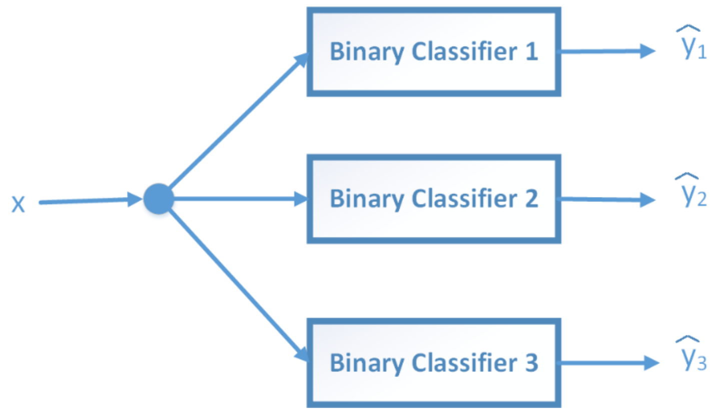
Show the following:

i. The design of the components of the classifier that predicts the category of a given measurement;
ii. How the data is presented to each component for training; and
iii. An example of how the class is inferred for a query data.

Answer.

(a) One-vs-all approach:

i. Since # classes (categories) is C = 4, we need C-1= 3 binary classifiers:




ii. The data is presented for training each classifier as follows:

| X | Target |
|---|---|
| 1 | 0 |
| 1.3 | 0 |
| 4.9 | **1** |
| 2.1 | 0 |
| 3.3 | 0 |
| 4.8 | **1** |
| 1.9 | 0 |
| 5.1 | **1** |

For training Binary Classifier 1
- Learns to classify 'A'

| X | Target |
|---|---|
| 1 | 0 |
| 1.3 | 0 |
| 4.9 | 0 |
| 2.1 | 0 |
| 3.3 | **1** |
| 4.8 | 0 |
| 1.9 | 0 |
| 5.1 | 0 |

For training Binary Classifier 2
- Learns to classify 'B'

| X | Target |
|---|---|
| 1 | 0 |
| 1.3 | 0 |
| 4.9 | 0 |
| 2.1 | **1** |
| 3.3 | 0 |
| 4.8 | 0 |
| 1.9 | **1** |
| 5.1 | 0 |

For training Binary Classifier 3
- Learns to classify 'C'

iii. To predict the category of a new measurement x, the predictions of the 3 classifiers are used:

$$\Pr(\text{category} = A) = \hat{y}_1$$

$$\Pr(\text{category} = B) = \hat{y}_2$$

$$\Pr(\text{category} = C) = \hat{y}_3$$

$$\Pr(\text{category} = D) = 1 - (\hat{y}_1 + \hat{y}_2 + \hat{y}_3)$$

The category with the highest probability above wins.

For example, assume that for x =5.0, we get the following output probabilities from the trained classifiers:

$$\Pr(\text{category} = A) = \hat{y}_1 = 0.4$$

$$\Pr(\text{category} = B) = \hat{y}_2 = 0.3$$

$$\Pr(\text{category} = C) = \hat{y}_3 = 0.2$$

We can calculate:

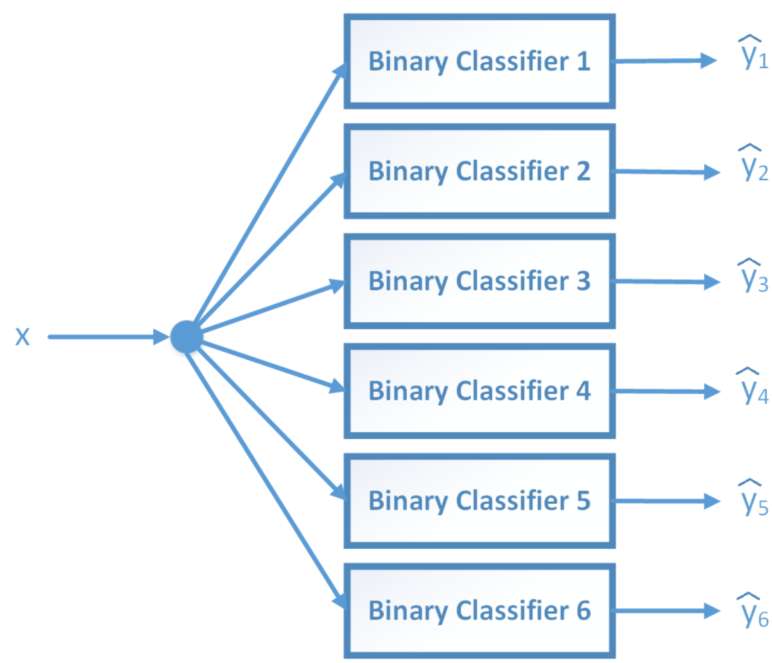
$$\Pr(\text{category} = D) = 1 - (\hat{y}_1 + \hat{y}_2 + \hat{y}_3) = 0.1$$

Therefore, since the highest probability is for category 'A', therefore the predicted category for x is 'A'.

**Note**: In the above example, the classifier outputs are probability values, i.e. each predictor's output is a value between 0 and 1 and the sum of all probabilities sums up to 1. If this is not the case, and the classifier outputs are not probability values, then we will need C=4 binary classifiers (one learning to classify each of the 4 categories). In that case, we simply predict the category corresponding with the highest classifier response.

(b) Pairwise approach:

i. Since # classes (categories) is C = 4, we need C (C-1)/2 = 6 binary classifiers:



ii. The data is presented for training each classifier as follows:

| X | Target |
|---|---|
| | |
| | |
| 4.9 | 1 |
| | |
| 3.3 | 0 |
| 4.8 | 1 |
| | |
| 5.1 | 1 |

For training Binary Classifier 1
- Learns to classify 'A' from 'B'

| X | Target |
|---|---|
| | |
| | |
| 4.9 | 1 |
| 2.1 | 0 |
| | |
| 4.8 | 1 |
| 1.9 | 0 |
| 5.1 | 1 |

For training Binary Classifier 2
- Learns to classify 'A' from 'C'

| X | Target |
|---|---|
| 1 | 0 |
| 1.3 | 0 |
| 4.9 | 1 |
| | |
| | |
| 4.8 | 1 |
| | |
| 5.1 | 1 |

For training Binary Classifier 3
- Learns to classify 'A' from 'D'

| X | Target |
|---|---|
| | |
| | |
| | |
| 2.1 | 0 |
| 3.3 | 1 |
| | |
| 1.9 | 0 |
| | |

For training Binary Classifier 4
- Learns to classify 'B' from 'C'

| X | Target |
|---|---|
| 1 | 0 |
| 1.3 | 0 |
| | |
| | |
| 3.3 | 1 |
| | |
| | |
| | |

For training Binary Classifier 5
- Learns to classify 'B' from 'D'

| X | Target |
|---|---|
| 1 | 0 |
| 1.3 | 0 |
| | |
| 2.1 | 1 |
| | |
| | |
| 1.9 | 1 |
| | |

For training Binary Classifier 6
- Learns to classify 'C' from 'D'

iii. To predict the category of a new measurement x, the predictions of the 6 classifiers are used:

Classifier 1: $Pr_{C1}(\text{category} = A) = \hat{y}_1$ $\quad Pr_{C1}(\text{category} = B) = 1 - \hat{y}_1$

Classifier 2: $Pr_{C2}(\text{category} = A) = \hat{y}_2$ $\quad Pr_{C2}(\text{category} = C) = 1 - \hat{y}_2$

Classifier 3: $Pr_{C3}(\text{category} = A) = \hat{y}_3$ $\quad Pr_{C3}(\text{category} = D) = 1 - \hat{y}_3$

Classifier 4: $Pr_{C4}(\text{category} = B) = \hat{y}_4$ $\quad Pr_{C4}(\text{category} = C) = 1 - \hat{y}_4$

Classifier 5: $Pr_{C5}(\text{category} = B) = \hat{y}_5$ $\quad Pr_{C5}(\text{category} = D) = 1 - \hat{y}_5$

Classifier 6: $Pr_{C6}(\text{category} = C) = \hat{y}_6$ $\quad Pr_{C6}(\text{category} = D) = 1 - \hat{y}_6$

For each category, take the average probability from above:

$$\Pr(\text{category} = A) = \frac{\hat{y}_1 + \hat{y}_2 + \hat{y}_3}{3}$$

$$\Pr(\text{category} = B) = \frac{(1-\hat{y}_1) + \hat{y}_4 + \hat{y}_5}{3}$$

$$\Pr(\text{category} = C) = \frac{(1-\hat{y}_2) + (1-\hat{y}_4) + \hat{y}_6}{3}$$

$$\Pr(\text{category} = D) = \frac{(1-\hat{y}_3) + (1-\hat{y}_5) + (1-\hat{y}_6)}{3}$$

The category with the highest average probability above wins.