CSCI-UA 472 Artificial Intelligence

Muhammad Wajahat Mirza

mwm356@nyu.edu

Homework 06

November 10, 2020

Problem 1

Suppose that we have the following data set:
Suppose that you are trying to carry out classification learning where M is the classification attribute and the rest are predictive attributes. You are given the following data set:

ID	A	B	D	M	Number of instances
1.	1	2	1	2	9
2.	1	2	1	3	22
3.	1	2	3	2	5
4.	1	2	2	1	4
5.	1	2	3	3	1
6.	2	1	1	1	1
7.	2	1	1	2	3
8.	2	1	1	3	2
9.	2	2	1	1	15
10.	2	2	2	3	6
11.	2	2	3	1	5
12.	2	2	3	3	16
13.	2	2	1	2	7
14.	2	2	1	3	3
15.	2	2	2	2	1

How does Naive Bayes classify the instance $A = 1, B = 1, D = 1$?
(Note that "number of instances" is not an attribute; you have to imagine that in the real data set there are 9 lines of type 1, with $A = 1, B = 2, D = 1, M = 2$, and so on.)

Solution to Problem 1

To do Naive Bayes classification of instance A = 1, B = 1, and D = 1, first estimate $P(M = v)$ as $Freq_T(M = v)$ where $v \in \{1, 2, 3\}$ and
$P(I_i|M = v)$ as $Freq_T(I_i = u|M = v)$ where $u = 1$ and I_i is each instance i.e. A, B, D.

Learning: Set $P(M = v)$ as $Freq_T(M = v)$ for all v in $Dom(M)$
Set $P(I_i|M = v)$ as $Freq_T(I_i = u|M = v)$ for $i = 1...k$, v in $Dom(M)$, $u = 1$ as asked in the question.

Inference: Compute $P(M = v) \times \prod_{i=1}^{k} P(I_i = u_i|M = v)$ for each value of v. We choose the value of v for which it is the largest.

- Let's find $P(M = v)$ for each v:
  - $P(M = v) = Freq_T(M = v) = \dfrac{(sum of instances|M = v)}{total}$ for given v
  - $Freq_T(M = 1) = \dfrac{25}{100} = 0.25$
  - $Freq_T(M = 2) = \dfrac{25}{100} = 0.25$
  - $Freq_T(M = 3) = \dfrac{50}{100} = 0.50$
- Let's find $P(I_i|M = v)$ for each I_i:
  - $P(I_i|M = v) = Freq_T(I_i = 1|M = v)$ for each I_i, v.

  - $Freq_T(A = 1|M = 1) = \dfrac{4}{25} = 0.16$
  - $Freq_T(A = 1|M = 2) = \dfrac{14}{25} = 0.56$
  - $Freq_T(A = 1|M = 3) = \dfrac{23}{50} = 0.46$

  - $Freq_T(B = 1|M = 1) = \dfrac{1}{25} = 0.04$
  - $Freq_T(B = 1|M = 2) = \dfrac{3}{25} = 0.12$
  - $Freq_T(B = 1|M = 3) = \dfrac{2}{50} = 0.04$

  - $Freq_T(D = 1|M = 1) = \dfrac{16}{25} = 0.64$

- $Freq_T(D = 1|M = 2) = \frac{19}{25} = 0.76$
- $Freq_T(D = 1|M = 3) = \frac{27}{50} = 0.54$

- So $P(M = 1|A = 1, B = 1, D = 1)$ is:
  - $\alpha \cdot P(M = 1) \cdot P(A = 1|M = 1) \cdot P(B = 1|M = 1) \cdot P(D = 1|M = 1)$
  - $\alpha \cdot 0.25 \cdot 0.16 \cdot 0.04 \cdot 0.64$
  - $0.001024\alpha \Rightarrow$ probability with M = 1.
- So $P(M = 2|A = 1, B = 1, D = 1)$ is:
  - $\alpha \cdot P(M = 2) \cdot P(A = 1|M = 2) \cdot P(B = 1|M = 2) \cdot P(D = 1|M = 2)$
  - $\alpha \cdot 0.25 \cdot 0.56 \cdot 0.12 \cdot 0.76$
  - $0.012768\alpha \Rightarrow$ probability with M = 2.
- So $P(M = 3|A = 1, B = 1, D = 1)$ is:
  - $\alpha \cdot P(M = 3) \cdot P(A = 1|M = 3) \cdot P(B = 1|M = 3) \cdot P(D = 1|M = 3)$
  - $\alpha \cdot 0.50 \cdot 0.46 \cdot 0.04 \cdot 0.54$
  - $0.004968\alpha \Rightarrow$ probability with M = 3.

Hence, the largest value acquired with given v values is $M = 2$ i.e. 0.012768α . Thus, for instances $A = 1, B = 1, D = 1$, $M = 2$ will be the value of M using Naive Bayes classification method.

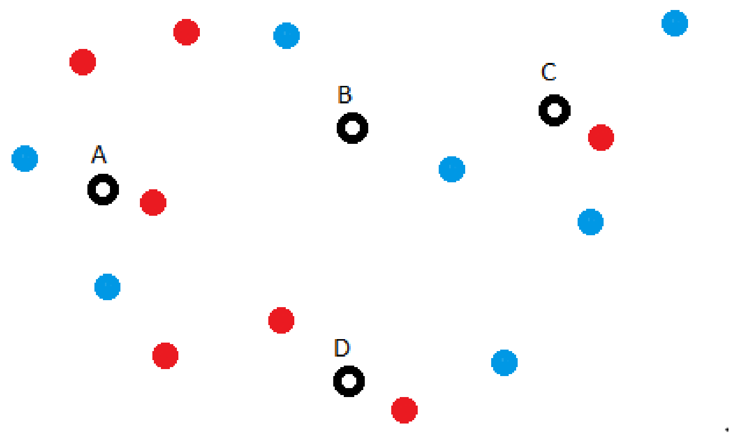
Problem 2

Suppose that your data is as pictured below. The predictive attributes are the x and y coordinates of the dots. The classification attribute is the color (red or blue). The colored points are the training data. The empty dots represent test points.

(A) How does 1-nearest neighbors classify the points A, B, C, and D? How does 3-nearest neighbors classify them?

(B) Suppose that you use the optimal linear separator. (Don't run a program; just experiment with a straight edge on the picture.) What is the maximum accuracy that you can obtain on the training set? How will it classify the points A, B, C, and D?

(C) Evaluate the linear separator that you found in (B) over the blue points in terms of recall, precision, and F-score.



Solution to Problem 2

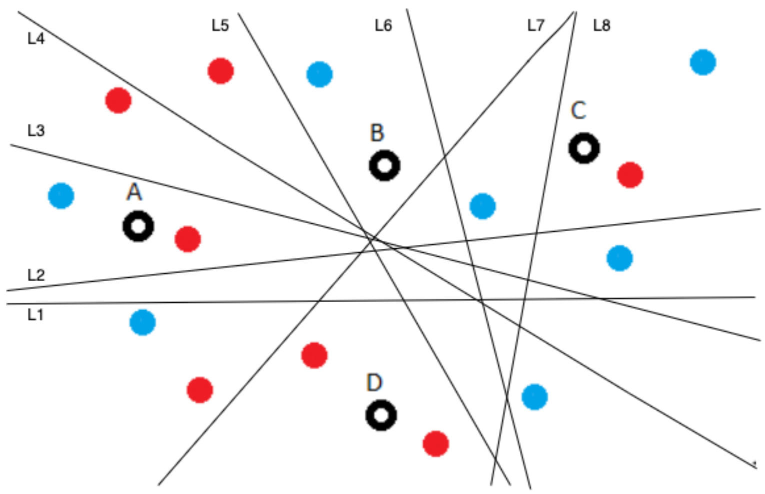
(A) Solution to A:

- **k-nearest neighbor (k-NN)**:
  As per **k-NN** classification, unknown test points such as A, B, C, D are classified on the basis of distance, either Euclidean distance $d^2(\vec{u}, \vec{v}) = \sum_{i=1}^{m}(\vec{u_i} - \vec{v_i})^2$ or Manhattan distance $d(\vec{u}, \vec{v}) = \sum_{i=1}^{m} |\vec{u_i} - \vec{v_i}|$.
- **1-nearest neighbor (1-NN)**:
  For **1-NN** classfication, "distance" is calculated between each of the training data point and the test data point. The point-set (train, test) with the smallest distance is categorized accordingly. For the given data, following would be classification based on 1-NN:
  - **A** will be classified as *Red* as it has the smallest distance to the red training data point.
  - **B** will be classified as *Blue* as it has the smallest distance to the blue training data point.
  - **C** will be classified as *Red* as it has the smallest distance to the red training data point.
  - **D** will be classified as *Red* as it has the smallest distance to the red training data point.
- **3-nearest neighbor (3-NN)**:
  For **3-NN** classfication, "distance" is calculated between each of the training data point and the test data point. The point-sets (train, test) with the majority of the smallest distance is categorized accordingly. For the given data, following would be classification based on 3-NN:
  - **A** will be classified as *Blue* as it has the smallest distance with 2 blue and 1 red training data points. Blue is the majority in 3-nearest-neighbor distance.
  - **B** will be classified as *Blue* as it has the smallest distance with 2 blue and 1 red training data points. Blue is the majority in 3-nearest-neighbor distance.

- **C** will be classified as *Blue* as it has the smallest distance with 2 blue and 1 red training data points. Blue is the majority in 3-nearest-neighbor distance.
- **D** will be classified as *Red* as it has the smallest distance with 2 red and 1 blue training data points. Red is the majority in 3-nearest-neighbor distance.

(B) Solution to B:

- To find optimal linear separator without running a program, let's try greedy method and find accuracy for different lines to get most optimal linear separator and accuracy. Some of the linear separators (L1, L2, ..., L8) tried are as in the following diagram:



- To find accuracy for each line, make confusion matrix and find accuracy using formula: $\dfrac{(TP + TN)}{Total}$ where TP = true positives, TN = true negatives.

- Confusion Matrix:

		prediction outcome	
		−Red	**+Blue**
actual value	**−Red**	True positive	False negative
	+Blue	False positive	True negative

- Use Confusion Matrix for each of the separator line, L1,...,L8 given in above diagram and calculate accuracy:

- **L1**

prediction outcome

		−Red	**+Blue**
actual value	**−Red**	3	4
	+Blue	2	5

Accuracy $= (3 + 5)/14 = 8/14 = 0.571$ i.e. 57.1%

- **L2**

prediction outcome

		−Red	**+Blue**
actual value	**−Red**	3	4
	+Blue	3	4

Accuracy $= (3 + 4)/14 = 7/14 = 0.5$ i.e. 50%

- **L3**

prediction outcome

		−Red	**+Blue**
actual value	**−Red**	4	3
	+Blue	3	4

Accuracy $= (4 + 4)/14 = 8/14 = 0.571$ i.e. 57.1%

- **L4**

prediction outcome

		−Red	**+Blue**
actual value	**−Red**	5	2
	+Blue	3	4

Accuracy $= (4+5)/14 = 9/14 = 0.643$ i.e. 64.3%

– **L5**

prediction outcome

		−Red	**+Blue**
actual value	**−Red**	6	1
	+Blue	2	5

Accuracy $= (6+5)/14 = 11/14 = 0.786$ i.e. 78.6%

– **L6**

prediction outcome

		−Red	**+Blue**
actual value	**−Red**	6	1
	+Blue	3	4

Accuracy $= (6+4)/14 = 10/14 = 0.714$ i.e. 71.4%

– **L7**

prediction outcome

		−Red	**+Blue**
actual value	**−Red**	4	3
	+Blue	3	4

Accuracy $= (4+4)/14 = 8/14 = 0.571$ i.e. 57.1%

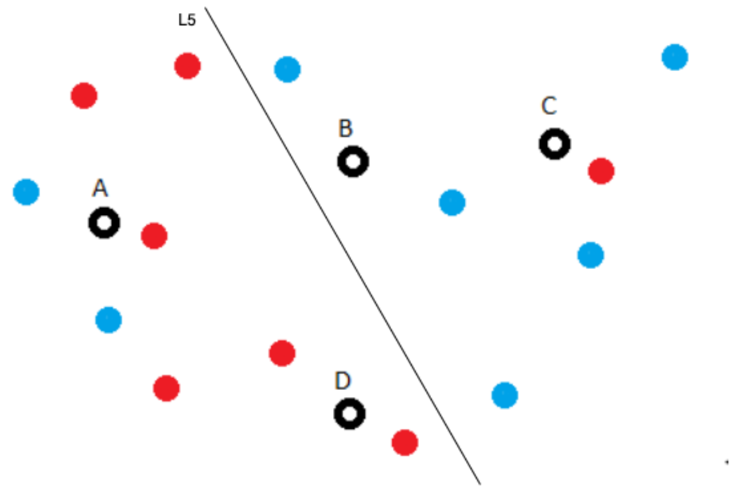
– **L8**

prediction outcome

		−Red	**+Blue**
actual value	**−Red**	6	1
	+Blue	4	3

Accuracy $= (6+3)/14 = 9/14 = 0.643$ i.e. 64.3%

- Solution remark: Hence, using the greedy method, we tried many linear separators to find an optimal line with the maximum accuracy. Maximum accuracy is achieved by L5. See the following diagram and confusion matrix to visualize optimal linear separator and the accuracy achieved by it. Thus, this linear separator categorizes **A, D** as Red and **B, C** as Blue.



- Confusion Matrix:

prediction outcome

		−Red	**+Blue**
actual value	**−Red**	6	1
	+Blue	2	5

Maximum Accuracy achieved is (6+5)/14 = 11/14 = 0.786 i.e. 78.6% $\Rightarrow$ Solution

(C) Solution to C, recall, precision and F-score over Blue points:

- To find Recall, use the following formula: $\dfrac{TP}{TP+FN}$ where TP = true positives, FN = False negatives.
  - Recall $= \dfrac{TP}{TP+FN} = \dfrac{5}{5+2} = \dfrac{5}{7} = 0.7143$
- To find Precision, use the following formula: $\dfrac{TP}{TP+FP}$ where TP = true positives, FN = False positives
  - Precision $= \dfrac{TP}{TP+FP} = \dfrac{5}{5+1} = \dfrac{5}{6} = 0.8333$
- To find F-score, use the following formula: $2 \times \dfrac{(Precision \times Recall)}{Precision + Recall}$
  - F-score $= 2 \times \dfrac{0.8333 \times 0.7143}{0.8333 + 0.7143} = 0.7692$

End of Assignment. Thank you!