ANS 1

(a)

List the points:

Index	(x, y)	Label
1	(1, 4)	–
2	(2, 4)	+
3	(1, 3)	–
4	(3, 3)	+
5	(5, 5)	–
6	(5, 1)	–

Now for each point, find the closest neighbor (Euclidean distance):
Point 1 → Closest to Point 2 (Distance = 1) → Point 2 label is + → Wrong
Point 2 → Closest to Point 1 (Distance = 1) → Point 1 label is − → Wrong
Point 3 → Closest to Point 1 (Distance = 1) → Point 1 label is − → Correct
Point 4 → Closest to Point 2 (Distance ≈ 1.41) and Point 3 (Distance ≈ 2.24) → Point 2 label is + → Correct
Point 5 → Closest to Point 4 (Distance ≈ 2.24) and others farther → Point 4 label is + → Wrong
Point 6 → Closest to Point 5 (Distance = 4) → Point 5 label is − → Correct

Wrong predictions = 3 (Points 1, 2, 5)
Total predictions = 6
Therefore, Error Rate = 3/6 = 0.5

(b)

Each point uses three nearest neighbors. If a tie happens, it's wrong.

Point 1: neighbors → (2,+), (3,–), (4,+) → 2+, 1− → Majority = + → Wrong
Point 2: neighbors → (1,–), (4,+), (3,–) → 1+, 2− → Majority = − → Wrong
Point 3: neighbors → (1,–), (2,+), (4,+) → 2+, 1− → Majority = + → Wrong
Point 4: neighbors → (2,+), (3,–), (5,–) → 1+, 2− → Majority = − → Wrong
Point 5: neighbors → (4,+), (6,–), (2,+) → 2+, 1− → Majority = + → Wrong
Point 6: neighbors → (5,–), (4,+), (2,+) → 2+, 1− → Majority = + → Wrong
All wrong predictions.

Therefore, Error Rate = 6/6 = 1.0

(c)

Distance-Weighted Voting: Closer neighbors have more weight (lower distance → higher influence).

Key: Weight = 1/d; which is Very close neighbors dominate.

<u>Point 1:</u>

(2,4) + → dist=1 → weight=1

(1,3) − → dist=1 → weight=1

(3,3) + → dist=√5≈2.24 → weight≈0.447

Weights: + (1 + 0.447) vs − (1) → + wins → Wrong

<u>Point 2:</u>

(1,4) − → dist=1 → weight=1

(3,3) + → dist=√2≈1.41 → weight≈0.707

(1,3) − → dist=√2≈1.41 → weight≈0.707

Weights: + (0.707) vs − (1+0.707) → − wins → Wrong

<u>Point 3:</u>

(1,4) − → dist=1 → weight=1

(2,4) + → dist=√2≈1.41 → weight≈0.707

(3,3) + → dist=2 → weight=0.5

Weights: + (0.707+0.5) vs − (1) → + wins → Wrong

<u>Point 4:</u>

(2,4) + → dist=√2≈1.41 → weight≈0.707

(1,3) − → dist=2 → weight=0.5

(5,5) − → dist=√5≈2.24 → weight≈0.447

Weights: + (0.707) vs − (0.5+0.447) → Tie → Wrong (Tie = Wrong)

<u>Point 5:</u>

(4,3) + → dist=√5≈2.24 → weight≈0.447

(5,1) − → dist=4 → weight=0.25

(2,4) + → dist=√10≈3.16 → weight≈0.316

Weights: + (0.447+0.316) vs − (0.25) → + wins → Wrong

<u>Point 6:</u>

(5,5) − → dist=4 → weight=0.25

(3,3) + → dist=√8≈2.83 → weight≈0.354

(2,4) + → dist=√17≈4.12 → weight≈0.243

Weights: + (0.354+0.243) vs − (0.25) → + wins → Wrong

All wrong predictions again.

Therefore, Error Rate = 1.0.

(d)

5					(-)	
4	(-)	(+)				
3	(-)		(+)			
2						
1					(-)	
0	0	1	2	3	4	5

x-axis →

ANS 2

(a)

Compute L1 distances:

L1 norm = sum of absolute differences.

<u>Distance to P1:</u>
|1−2|+|2−3|+|3−4|+|4−5|+|4−5|=1+1+1+1+1=5

<u>Distance to P2:</u>
|1−0|+|2−1|+|3−2|+|4−3|+|4−5|=1+1+1+1+1=5|

<u>Distance to P3:</u>
|1−2|+|2−2|+|3−2|+|4−2|+|4−4|=1+0+1+2+0=4

<u>Distance to P4:</u>
|1−3|+|2−3|+|3−3|+|4−3|+|4−3|=2+1+0+1+1=5

<u>Distance to P5:</u>
|1−4|+|2−2|+|3−4|+|4−4|+|4−4|=3+0+1+0+0=4

Rank distances

Point	Distance	Class
P3	4	Sell
P5	4	Sell
P1	5	Buy
P2	5	Buy
P4	5	Sell

Top 3 nearest neighbors: P3 (Sell), P5 (Sell), P1 (Buy).

Majority vote

- Sell: 2 votes (P3, P5)
- Buy: 1 vote (P1)

Therefore, Final Prediction = Sell

ANS 4

Step 1: Calculate class priors

$P(+) = 5 \div 10 = 0.5$

$P(-) = 5 \div 10 = 0.5$

Step 2: Conditional probabilities using m-estimate ($m = 4$, $p = 0.5$)

For class +:

$P(A=0 \mid +) = (2 + 4(0.5)) \div (5 + 4) = 4 \div 9$

$P(B=1 \mid +) = (2 + 4(0.5)) \div (5 + 4) = 4 \div 9$

$P(C=0 \mid +) = (2 + 4(0.5)) \div (5 + 4) = 4 \div 9$

For class -:

$P(A=0 \mid -) = (3 + 4(0.5)) \div (5 + 4) = 5 \div 9$

$P(B=1 \mid -) = (2 + 4(0.5)) \div (5 + 4) = 4 \div 9$

$P(C=0 \mid -) = (1 + 4(0.5)) \div (5 + 4) = 3 \div 9 = 1 \div 3$

Step 3: Calculate final scores

For +:

Score(+) = P(+) × P(A=0 | +) × P(B=1 | +) × P(C=0 | +)

= 0.5 × (4 ÷ 9) × (4 ÷ 9) × (4 ÷ 9)

= 0.5 × (64 ÷ 729)

= 32 ÷ 729 ≈ 0.0439

For -:

Score(-) = P(-) × P(A=0 | -) × P(B=1 | -) × P(C=0 | -)

= 0.5 × (5 ÷ 9) × (4 ÷ 9) × (1 ÷ 3)

= 0.5 × (20 ÷ 243)

= 10 ÷ 243 ≈ 0.0412

Step 4: Conclusion

Since

Score(+) > Score(-)

The Naïve Bayes classifier will predict + for (A=0, B=1, C=0)

ANS 5

(a)

Learning each CPT from the dataset

Pollution (P):

P(Low) = (count of L) ÷ (total examples)

P(High) = (count of H) ÷ (total examples)

Smoker (S):

P(S=TRUE) = (count of TRUE) ÷ (total examples)

P(S=FALSE) = (count of FALSE) ÷ (total examples)

Cancer (C | P, S):

For each combination (P,L)(S,TRUE), find:

P(C=TRUE | P=x, S=y)

XRay (X | C):

P(X=pos | C=TRUE)

P(X=pos | C=FALSE)

Dyspnoea (D | C):

P(D=TRUE | C=TRUE)

P(D=TRUE | C=FALSE)

Learning conditional probabilities

Pollution:
Total = 20 examples

P(L) = 12 ÷ 20 = 0.6

P(H) = 8 ÷ 20 = 0.4

Smoker:
S=TRUE: 9 examples

S=FALSE: 11 examples

Thus:

P(S=TRUE) = 9 ÷ 20 = 0.45

P(S=FALSE) = 11 ÷ 20 = 0.55

Cancer (C | P, S):

When Pollution = Low (L) and Smoker = True:

P(C=TRUE | P=L, S=TRUE) = 1 ÷ 4 = 0.25

P(C=FALSE | P=L, S=TRUE) = 3 ÷ 4 = 0.75

When Pollution = Low (L) and Smoker = False:

P(C=TRUE | P=L, S=FALSE) = 1 ÷ 8 = 0.125

P(C=FALSE | P=L, S=FALSE) = 7 ÷ 8 = 0.875

When Pollution = High (H) and Smoker = True:

P(C=TRUE | P=H, S=TRUE) = 4 ÷ 5 = 0.8

P(C=FALSE | P=H, S=TRUE) = 1 ÷ 5 = 0.2

When Pollution = High (H) and Smoker = False:

P(C=TRUE | P=H, S=FALSE) = 1 ÷ 3 ≈ 0.333

P(C=FALSE | P=H, S=FALSE) = 2 ÷ 3 ≈ 0.666

XRay (X | C):
C=TRUE:

6 examples

X=pos happens 6 times

C=FALSE:

14 examples

X=pos happens 3 times

Thus:

P(X=pos | C=TRUE) = 6 ÷ 6 = 1.0

P(X=pos | C=FALSE) = 3 ÷ 14 ≈ 0.214

Dyspnoea (D | C):

C=TRUE:

D=TRUE occurs 4 times out of 6

C=FALSE:

D=TRUE occurs 4 times out of 14

Thus:

P(D=TRUE | C=TRUE) = 4 ÷ 6 ≈ 0.667

P(D=TRUE | C=FALSE) = 4 ÷ 14 ≈ 0.286

(b)

Compute P(D|E)

Evidence: E = P = L, S = TRUE, C = TRUE, X = pos

We compute:

P(D = TRUE | E) and P(D = FALSE | E)

First, unnormalized joint probability:

P(D, E) = P(P = L) × P(S = TRUE) × P(C = TRUE | P, S) × P(X = pos | C) × P(D | C)

Calculate:

For D = TRUE:

P(D = TRUE, E) = 0.6 × 0.45 × 0.25 × 1.0 × 0.667

P(D = TRUE, E) = 0.045

For D = FALSE:

Remember:

P(D = FALSE | C = TRUE) = 1 − P(D = TRUE | C = TRUE)

P(D = FALSE | C = TRUE) = 1 − 0.667 = 0.333

Thus:

P(D = FALSE, E) = 0.6 × 0.45 × 0.25 × 1.0 × 0.333

P(D = FALSE, E) = 0.0225

<u>Normalize</u>

Sum:

P(D = TRUE, E) + P(D = FALSE, E) = 0.045 + 0.0225 = 0.0675

Thus:

P(D = TRUE | E) = 0.045 ÷ 0.0675 ≈ 0.6667

P(D = FALSE | E) = 0.0225 ÷ 0.0675 ≈ 0.3333

ANS 3 & 6

Link to Github