CPSC 340: Assignment 1

Parker Ram (z1l1b)

43156165

Problem 1

1.1

1.) 14

2.) 0

3.) (2,6,2)

4.) $\sqrt{5}$

5.) (6,5,7)

6.) 19

7.) $\begin{bmatrix} 11 & 10 & 10 \\ 10 & 14 & 10 \\ 10 & 10 & 14 \end{bmatrix}$

1.2 1.)True 2.)True 3.)True 4.)True 5.)False 6.)True 7.)False 8.)False 9.)True 10.)False

Problem 2

2.1

1.) $5

2.) 0.55

3.) 0.92

2.2

1.) 0.0101

2.) Most of the positive tests come from false positives

3.) 0.0096

4.) If this test was given to a random person and came back positive, it is not likely they are actually a drug user.

5.) Increasing $P(T = 0|D = 0)$ will make the test more useful.

Problem 3

3.1

1.) $6x - 2$

2.) $1 - 2x$

3.) $1 - \frac{e^{-x}}{p(x)}$

3.2

1.) $(2x_1 + e^{x_1+2x_2}, 2e^{x_1+2x_2})$

2.) $\frac{1}{Z}(e^{x_1}, e^{x_2}, e^{x_3})$

3.) a^T

4.) $(2x_1 - x_2, -x_1 + 2x_2)$

5.) x

3.3

1.) $\frac{14}{3}$

2.) $\frac{1}{4}$

3.) 0

4.) 0.5

5.) 1

6.) 0

Problem 4

4.1

1.) 6

2.) 6

4.2

1.) O(n)

2.) O(logn)

3.) Avg: O(1) Worst: O(n)

4.) O(d)

5.) O(d^2)

4.3

1.) O(N)

2.) O(N)

3.) O(1)

4.) O(N^2)

Problem 5

5.1

1.) Min: 0.352 — Max: 4.862 — Mean: 1.325 — Median: 1.159 — Mode: 0.77

2.) 5%: 0.465 — 25%: 0.718 — 50%: 1.159 — 75%: 1.813 — 95%: 2.624

5.2

1.) Plot D: If showing distribution of EACH going to need legend

2.) Plot C: Simple histogram

3.) Plot B: X-axis is weeks

4.) Plot A: Y-axis is illness percentage and is line graph

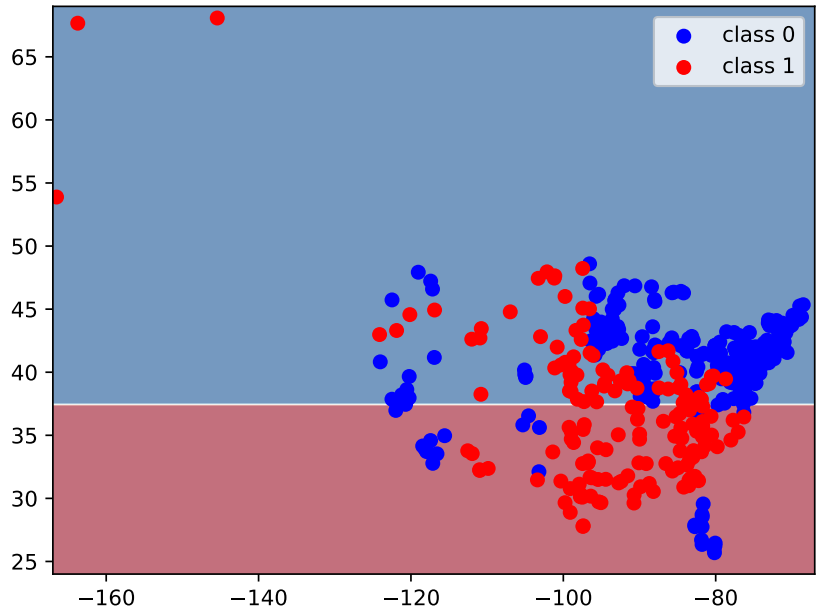
5.) Plot F: Dots very close together

6.) Plot E: Dots not as close together

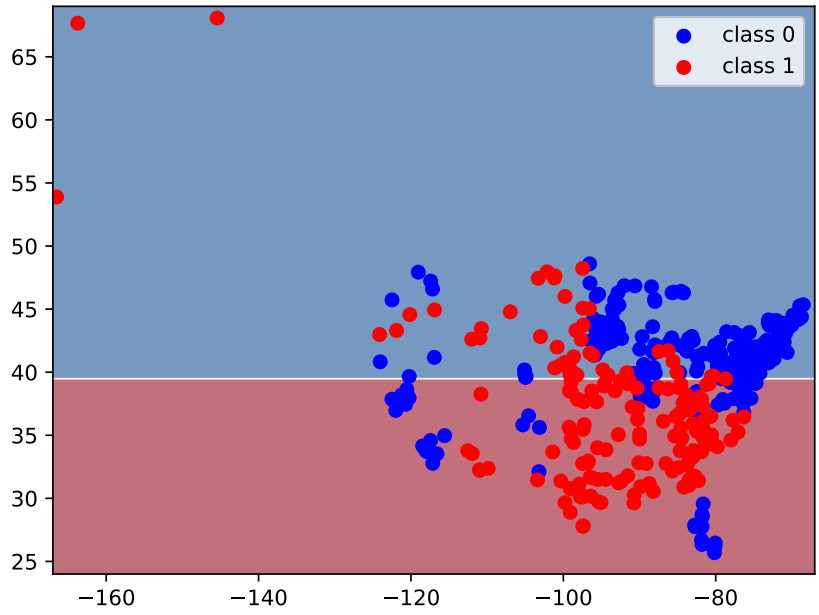
Problem 6

6.1) If the feature is binary or consists of a small set of discrete values it makes sense to use equality-based splitting rules because it can only take on a small number of possibilities.

6.2) Error: 0.265



6.3) Error: 0.275



6.5)

6.6)

6.7)