# Assign3\_naive\_bayes\_classifier

# James Bao (PSU ID: 934097519)

1)

You are a meteorologist that places temperature sensors all of the world, and you set them up so that they automatically e-mail you, each day, the high temperature for that day. Unfortunately, you have forgotten whether you placed a certain sensor S in Portland or in the Sahara desert (but you are sure you placed it in one of those two places) . The probability that you placed sensor S in Portland is 5%. The probability of getting a daily high temperature of 80 degrees or more is 20% in Portland and 90% in Sahara. Assume that probability of a daily high for any day is conditionally independent of the daily high for the previous day, given the location of the sensor.

**Part a:** If the first e-mail you got from sensor S indicates a daily high under 80 degrees, what is the probability that the sensor is placed in Portland?

**Part b:** If the first e-mail you got from sensor S indicates a daily high under 80 degrees, what is the probability that the second e-mail also indicates a daily high under 80 degrees?

**Part c:** What is the probability that the first three e-mails all indicate daily highs under 80 degrees?

a. P(<80 F | PDX) = 1.00 – 0.20 = 0.80. P(<80 F | Sahara) = 1.00 – 0.90 = 0.10

P(PDX) = 0.05, P(Sahara) = 0.95.

P(<80 F | PDX) \* P(PDX) = 0.80 \* 0.05 = 0.04

P(<80 F | Sahara) \* P(Sahara) = 0.10 \* 0.95 = 0.095

P(PDX) = P(<80 F | PDX) \* P(PDX)/( P(<80 F | PDX) \* P(PDX) + P(<80 F | Sahara) \* P(Sahara)) = 0.04 / (0.04 + 0.095) = **0.2963.**

b. P(<80 F) = P(<80 F | PDX) \* P(PDX) + P(<80 F | Sahara) \* P(Sahara) = 0.04 + 0.095 = **0.135.**

c. Assuming all 3 events are independent, P(<80 F for 3 days) = 0.135 \* 0.135 \* 0.135 = **2.46 \* 10-3.**

2)

Function P is a function defined on a set of samples S = {A, B, C, D}. We do not know the value of P for all samples, but we know that P(A) = 0.3 and P(B) = 0.6. What can you say about whether P is a valid probability function? Is P definitely a probability function, possibly a probability function, or definitely not a probability function? Justify your answer.

P is possibly a probability function. The criteria for a probability function are: 1) each value for every event defined by the function must be between 0 and 1, and 2) the sum of the probabilities of all events must be equal to 1. P(A) = 0.3 and P(B) = 0.6, so if P(C) + P(D) = 0.1, then function P is a probability function.

3)

Function P is a function defined on the set of real numbers. We do not know the value of P for all cases, but we know that P(x) = 0.3 when 0 <= x <= 10. What can you say about whether P is a valid probability density function? Is P definitely a probability density function, possibly a probability density function, or definitely not a probability density function? Justify your answer.

P is definitely not a probability function in this case. If P(x) has a constant value of 0.3 across the range 0 <= x <= 10, then the integral of P(x) across that function is 3.0, which is greater than 1. A probability function has an integral across all real numbers that is equal to 1.

4) NOTE: the program output uses the log\_e approximation for classÎ{+1,-1} argmax P(class) P(xi i Õ |class) in the Naïve Bayes classifier

Training data (yeast\_training.txt)-

Class 1 , attribute 1 , mean = 0.52 , std = 0.10

Class 1 , attribute 2 , mean = 0.54 , std = 0.10

Class 1 , attribute 3 , mean = 0.52 , std = 0.07

Class 1 , attribute 4 , mean = 0.41 , std = 0.17

Class 1 , attribute 5 , mean = 0.50 , std = 0.01

Class 1 , attribute 6 , mean = 0.00 , std = 0.01

Class 1 , attribute 7 , mean = 0.50 , std = 0.05

Class 1 , attribute 8 , mean = 0.24 , std = 0.05

Class 2 , attribute 1 , mean = 0.45 , std = 0.11

Class 2 , attribute 2 , mean = 0.45 , std = 0.10

Class 2 , attribute 3 , mean = 0.53 , std = 0.06

Class 2 , attribute 4 , mean = 0.23 , std = 0.11

Class 2 , attribute 5 , mean = 0.50 , std = 0.04

Class 2 , attribute 6 , mean = 0.00 , std = 0.01

Class 2 , attribute 7 , mean = 0.49 , std = 0.06

Class 2 , attribute 8 , mean = 0.33 , std = 0.14

Class 3 , attribute 1 , mean = 0.43 , std = 0.10

Class 3 , attribute 2 , mean = 0.48 , std = 0.11

Class 3 , attribute 3 , mean = 0.36 , std = 0.06

Class 3 , attribute 4 , mean = 0.22 , std = 0.08

Class 3 , attribute 5 , mean = 0.51 , std = 0.05

Class 3 , attribute 6 , mean = 0.00 , std = 0.01

Class 3 , attribute 7 , mean = 0.51 , std = 0.04

Class 3 , attribute 8 , mean = 0.27 , std = 0.09

Class 4 , attribute 1 , mean = 0.79 , std = 0.07

Class 4 , attribute 2 , mean = 0.76 , std = 0.07

Class 4 , attribute 3 , mean = 0.38 , std = 0.06

Class 4 , attribute 4 , mean = 0.32 , std = 0.11

Class 4 , attribute 5 , mean = 0.50 , std = 0.01

Class 4 , attribute 6 , mean = 0.00 , std = 0.01

Class 4 , attribute 7 , mean = 0.51 , std = 0.07

Class 4 , attribute 8 , mean = 0.27 , std = 0.09

Class 5 , attribute 1 , mean = 0.74 , std = 0.15

Class 5 , attribute 2 , mean = 0.62 , std = 0.12

Class 5 , attribute 3 , mean = 0.42 , std = 0.08

Class 5 , attribute 4 , mean = 0.30 , std = 0.12

Class 5 , attribute 5 , mean = 0.50 , std = 0.01

Class 5 , attribute 6 , mean = 0.00 , std = 0.01

Class 5 , attribute 7 , mean = 0.51 , std = 0.06

Class 5 , attribute 8 , mean = 0.24 , std = 0.04

Class 6 , attribute 1 , mean = 0.54 , std = 0.14

Class 6 , attribute 2 , mean = 0.50 , std = 0.12

Class 6 , attribute 3 , mean = 0.51 , std = 0.05

Class 6 , attribute 4 , mean = 0.24 , std = 0.10

Class 6 , attribute 5 , mean = 0.50 , std = 0.01

Class 6 , attribute 6 , mean = 0.49 , std = 0.38

Class 6 , attribute 7 , mean = 0.51 , std = 0.03

Class 6 , attribute 8 , mean = 0.24 , std = 0.05

Class 7 , attribute 1 , mean = 0.48 , std = 0.11

Class 7 , attribute 2 , mean = 0.47 , std = 0.09

Class 7 , attribute 3 , mean = 0.54 , std = 0.06

Class 7 , attribute 4 , mean = 0.22 , std = 0.12

Class 7 , attribute 5 , mean = 0.50 , std = 0.04

Class 7 , attribute 6 , mean = 0.00 , std = 0.03

Class 7 , attribute 7 , mean = 0.50 , std = 0.06

Class 7 , attribute 8 , mean = 0.26 , std = 0.09

Class 8 , attribute 1 , mean = 0.74 , std = 0.10

Class 8 , attribute 2 , mean = 0.73 , std = 0.11

Class 8 , attribute 3 , mean = 0.49 , std = 0.05

Class 8 , attribute 4 , mean = 0.29 , std = 0.07

Class 8 , attribute 5 , mean = 0.50 , std = 0.01

Class 8 , attribute 6 , mean = 0.00 , std = 0.01

Class 8 , attribute 7 , mean = 0.46 , std = 0.08

Class 8 , attribute 8 , mean = 0.23 , std = 0.02

Class 9 , attribute 1 , mean = 0.55 , std = 0.14

Class 9 , attribute 2 , mean = 0.56 , std = 0.15

Class 9 , attribute 3 , mean = 0.51 , std = 0.06

Class 9 , attribute 4 , mean = 0.20 , std = 0.06

Class 9 , attribute 5 , mean = 0.50 , std = 0.01

Class 9 , attribute 6 , mean = 0.00 , std = 0.01

Class 9 , attribute 7 , mean = 0.53 , std = 0.05

Class 9 , attribute 8 , mean = 0.24 , std = 0.04

Class 10 , attribute 1 , mean = 0.78 , std = 0.05

Class 10 , attribute 2 , mean = 0.73 , std = 0.11

Class 10 , attribute 3 , mean = 0.48 , std = 0.09

Class 10 , attribute 4 , mean = 0.33 , std = 0.06

Class 10 , attribute 5 , mean = 1.00 , std = 0.01

Class 10 , attribute 6 , mean = 0.00 , std = 0.01

Class 10 , attribute 7 , mean = 0.55 , std = 0.02

Class 10 , attribute 8 , mean = 0.23 , std = 0.01

Classification (yeast\_test.txt)-

classification accuracy = 0.3430