Bayes Classifier

Problem 1

Assume that you have data from $\mathbb{R}^2$, which belong to classes ω_0 or ω_1. These classes have the same prior probability. The probability distribution for data from class ω_0 is uniform in the circle defined by $\|\mathbf{x}\| < 4$, whereas the probability distribution for data from class ω_1 is Gaussian, given by

$$p(\mathbf{x}|\omega_1) = \frac{1}{2\pi} e^{\frac{-\|\mathbf{x}\|^2}{2}}.$$

a) Find the classification boundary of the Bayes Classifier for these data. Identify which regions of 2D space are classified into ω_0 and which are classified into ω_1.

b) Find the probability that data from class ω_0 are classified into ω_1.

Problem 2

Assume that you are using a naive Bayes classifier to recognize the languages in which pieces of text are written. The classifier uses trigrams (three-character sequences) as features. The numbers of occurrences of trigrams in the training data are as follows, with blank spaces represented by '*b*':

Trigram	Portuguese	French	English
*b*eu	8	6	1
cho	9	3	1
e*b*c	6	4	5
eu*b*	9	3	2
hoj	7	1	2
hov	9	2	3
je*b*	1	4	1
oje	6	1	2
ove	5	2	3
all	1	7	10

Find in which language the classifier places the sentence "hoje chove". Assume that the prior probabilities for all languages are equal and that the number of different trigrams is 27^3. Use Laplace smoothing.