Homework III

1. (Point Estimation) You are given a coin and a thumbtack and you put Beta priors Beta(100; 100) and Beta(1; 1) on the coin and thumbtack respectively. You perform the following experiment: toss both the thumbtack and the coin 100 times. To your surprise, you get 60 heads and 40 tails for both the coin and the thumbtack. Are the following two statements true or false?
   _ The MLE estimate of both the coin and the thumbtack is the same but the MAP estimate is not.
   _ The MAP estimate of the parameter θ (probability of landing heads) for the coin is greater than the MAP estimate of θ for the thumbtack.
   Explain your answer mathematically. **[5 Points]**

2. Point Estimation
   Given that it is virtually impossible to find a suitable "date" for boring, geeky computer scientists, you start a dating website called "www.csdating.com." Before you launch the website, you do some tests in which you are interested in estimating the failure probability of a "potential date" that your website recommends. In order to do that, you perform a series on experiments on your classmates (friends). You ask them to go on "dates" until they find a suitable match. The number of failed dates, k, is recorded.

   a. Given that p is the failure probability, what is the probability of k failures before a suitable "match" is found by your friend.
   b. You have performed m independent experiments of this form (namely, asked m of your friends to go out on dates until they find a suitable match), recording k_1,........ k_m. Estimate the most likely value of p as a function of m and k_1,........ k_m.

   [8 Points]

3. Naive Bayes is a linear classifier. True or False. Explain. **[2 Points]**

4. Consider learning a function X -> Y where Y is boolean, where X $\langle X_1, X_2\rangle$, and where X_1 is a boolean variable and X_2 is continuous variable. State the parameters that must be estimated to define a Naïve Bayes classifier in this case. Give the formula for computing P(Y|X), in terms of these parameters and the feature values X_1 and X_2.

   [5 Points]

5. CLASSIFICATION :
   Imagine that you are given the following set of training examples. Each feature can take on one of three nominal values: a, b, or c.

F1	F2	F3	Category
a	c	a	+
c	a	c	+
a	a	c	-
b	c	a	-
c	c	b	-

How would a Naive Bayes system classify the following test example? Be sure to show your work.

F1 = a, F2 = c , F3 = b

[10 Points]

6. Naïve Bayes
   Classify whether a given person is a male or a female based on the measured features using naïve bayes classifier. The features include height, weight, and foot size.

[10 Points]

Training Data for the classifier is given in the below table.

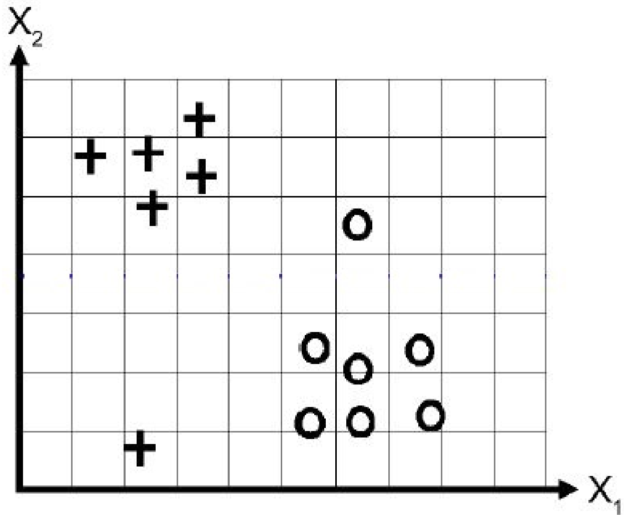
Person	height (feet)	weight (lbs)	foot size(inches)
male	6	180	12
male	5.92 (5'11")	190	11
male	5.58 (5'7")	170	12
male	5.92 (5'11")	165	10
female	5	100	6
female	5.5 (5'6")	150	8
female	5.42 (5'5")	130	7
female	5.75 (5'9")	150	9

Below is a sample to be classified as male or female.

Person	height (feet)	weight (lbs)	foot size(inches)
sample	6	130	8

7. Regularization separate terms in 2d logistic regression

[10 Points]



a. Consider the data in Figure where we fir model $p(y = 1 \mid x, w) = \sigma(w_0 + w_1 x_1 + w_2 x_2)$. Suppose we fit the model by maximum likelihood or we minimize
$J(w) = -l(w, D_{train})$

where $l(w, D_{train})$ is the log likelihood on the training set. Sketch a possible decision boundary corresponding to w. Is your answer (decision boundary) unique? How many classification errors does your method make on the training set?

b. Now suppose we regularize only the w_0 parameter i.e. we minimize

$$J(w) = -l(w, D_{train}) + \lambda w_0^2$$

Suppose λ is a very large number, so we regularize w_0 all the way to 0, but all other parameters are unregularized. Sketch a possible decision boundary. How many classification errors does your method make on training set?

c. Now suppose we regularize only the w_1 parameter i.e. we minimize
$J(w) = -l(w, D_{train}) + \lambda w_1^2$
Sketch a possible decision boundary. How many classification errors does your method make on training set?

d. Now suppose we regularize only the w_2 parameter i.e. we minimize
$J(w) = -l(w, D_{train}) + \lambda w_2^2$
Sketch a possible decision boundary. How many classification errors does your method make on training set?