1. (Naive Bayes, 10 points) Show that Naive Bayes classifier is a linear classifier in the following sense. Suppose an instance $x_i = (x_{i1}, \ldots, x_{id})$ has $d$ binary features. Then the classification rule of a Naive Bayes classifier is actually

$$\hat{y}_i = \begin{cases} 1 & \text{if } g_i > 0 \\ 0 & \text{otherwise} \end{cases},$$

with $g_i = \alpha_0 + \sum_{j=1}^{d} \alpha_j(x_i)$. To understand the decision rule as a linear classifier, we can think that Naive Bayes classifier first use functions, $\alpha_j(\cdot), j = 0, \ldots, d$, to transform the input feature first and then try to separate positives from negatives. Suppose the classifier fits these functions from a training set with data matrix $X$. Can you find out the form of these functions $\alpha_j(\cdot), j = 0, \ldots, d$?

2. (Linear classifier, 3 points) Suppose we have $d + 1$ training instances, $X = (x_1, \ldots, x_{(d+1)})$ with each instance $x_i$ has $d$ features. These instances have binary labels in a vector $y = (y_1, \ldots, y_{(d+1)})$. We further assume that any $2 \leq k \leq d + 1$, instances are not on a $k - 2$ dimensional hyperplane. Show that there is a linear classifier that can perfectly separate positive training instances from negative ones. (Hint: use induction. If you cannot work out details of the induction step, just write down your thoughts.)

3. (Optimization 7 points) Suppose we have function $f(x) = \begin{cases} -x & \text{if } x \leq 0 \\ x & \text{if } x > 0 \end{cases}$

Now we run the gradient descent procedure to solve this minimization problem,

$$\min_x f(x)$$

Suppose our start point $x_0 = 1.1$, and our step length is fixed at 0.2. Answer the following questions.

1) Where is the minimum of this function?

2) Does this gradient descent procedure converge? Why?

3) Can you find values for $x_0$ such that the gradient descent procedure converges?