Practice Problem Set 7 - Optimization, LDA, Naive Bayes

1. (Chang, NCU Taiwan) Consider the optimization problem

$$\min ||Ax - b||^2,$$

where $A \in \Re^{m \times n}, m \geq n, b \in \Re^m$. Compute the gradient and Hessian of the objective function. Write down the fixed step size gradient descent algorithm for solving the optimization problem.

2. (Pena, CMU) Prove that for a quadratic cost function, Newton's method reaches the optimum point in one step.

3. (Adapted from Boyd) Consider the quadratic objective function $f(x) = \frac{1}{2}(x_1^2 + \gamma x_2^2)$, where $\gamma > 0$. Consider the initial starting point $x^{(0)} = \begin{bmatrix} \gamma & 1 \end{bmatrix}^\top$. Derive the steepest descent update equation, and show the general term of $x^{(k)}$ for $k = 0, 1, 2, \cdots$.

4. (Boyd) Consider the following problem with $x \in \Re^2$.

$$\begin{array}{rl} \text{minimize} & x_1^2 + x_2^2 \\ \text{subject to} & (x_1 - 1)^2 + (x_2 - 1)^2 \leq 1 \\ \text{subject to} & (x_1 - 1)^2 + (x_2 + 1)^2 \leq 1 \end{array}$$

   (a) Sketch the feasible set and level sets of the objective. Find the optimal point x^* and optimal value p^*.

   (b) Write the KKT conditions and the Lagrangian.

5. Murphy 9.1

6. (StatQuest) Consider that there are two types of messages, normal and spam. Normal messages have the following words that occur: "Dear" 8 times, "Friend" 5 times, "Lunch" 3 times, and "Money" 1 time. Spam messages have words occurring with different frequencies: "Dear" 2 times, "Friend" 1 time, "Lunch" zero times, and "Money" 4 times.

   Now consider a new message with the words "Dear Friend". Consider the words are independent of each other. The prior probability that any message is a normal message is $2/3 (\approx 0.67)$.

   (a) Using Naive Bayes, is the new message normal or spam?

   (b) Does it matter if the message was "Friend Dear"? What does that say about the independence assumption?

   (c) What would happen if one of the words was "Lunch"?