Math 163 Introductory Seminar - Lehigh University - Spring 2008 - Assignment 7 Solutions
Due Wednesday March 19

24. Attend at least one of the lectures by Persi Diaconis and write a short paragraph describing what you thought of it. Include at least one simple idea presented so it is clear that you were really there. You can turn this in Friday March 21.

25. Consider the system of equations (in one variable) $x = a_1, x = a_2, \ldots, x = a_n$ for given numbers $a_1 \leq a_2 \leq \cdots \leq a_n$. The best L_1 approximation is any value of x that minimizes $\sum_{i=1}^{n} |a_i - x|$. Prove that the median is such a value. For simplicity assume that $n = 2m+1$ is odd and hence the median is a_{m+1}.
Hints - Use $|a_i - x| \geq a_i - x$, $|a_i - x| \geq -(a_i - x) = x - a_i$ and $|a_i - x| \geq 0$. Apply these bounds to the terms in $\sum_{i=1}^{n} |a_i - x|$ to show that it is always at least the value you get for the median. Consider which bounds you may want to apply to terms for $i < m+1$, $i > m+1$ and $i = m+1$ separately.

For any x we have $\sum_{i=1}^{n} |a_i - x| = (\sum_{i=1}^{m} |a_i - x|) + |a_{m+1} - x| + (\sum_{i=m+2}^{n} |a_i - x|) \geq (\sum_{i=1}^{m} x - a_i) + 0 + (\sum_{i=m+2}^{n} a_i - x) = \sum_{i=1}^{m} -a_i + \sum_{i=m=2}^{n} a_i$. When $x = a_{m+1}$ each of these is equality so we get $\sum_{i=1}^{n} |a_i - x| \geq \sum_{i=1}^{n} |a_i - a_{m+1}| = \sum_{i=1}^{m} -a_i + \sum_{i=m=2}^{n} a_i$ and the median is a minimizer.

26. Consider the system of equations (in one variable) $x = a_1, x = a_2, \ldots, x = a_n$ for given numbers $a_1 \leq a_2 \leq \cdots \leq a_n$. The best L_2 approximation is any value of x that minimizes $\sum_{i=1}^{n} (a_i - x)^2$. Prove that this is the arithmetic mean $\frac{1}{n} \sum_{i=1}^{n} a_i$.
Hint - use elementary calculus.

This is a function in the single variable x. Differentiating with respect to x and setting the result to 0 we get $0 = \sum_{i=1}^{n} -2(a_i - x) = -2\sum_{i=1}^{n} a_i + 2\sum_{i=1}^{n} x = -2\sum_{i=1}^{n} a_i + nx$ and hence $x = \frac{1}{n} \sum_{i=1}^{n} a_i$ is a critical point. The second derivative is $n > 0$ and so it minimizes.

27. Consider the system of equations (in one variable) $x = a_1, x = a_2, \ldots, x = a_n$ for given numbers $a_1 \leq a_2 \leq \cdots \leq a_n$. The best L_∞ approximation is any value of x that minimizes $\max_i\{|a_i - x|\}$. Prove that this is the midrange $\frac{a_1 + a_n}{2}$.
Hint - This is very straightforward. Note that $\max_i\{|a_i - x|\}$ is at least the value you get for the midrange when x is at most the midrange and when x is at least the midrange.

When x is at most the midrange, $x \leq \frac{a_1 + a_n}{2} \leq a_n$ and then $\max_i\{|a_i - x|\} \geq (a_n - x) \geq a_n - \frac{a_1 + a_n}{2} = \frac{a_n - a_1}{2}$. When x is at least the midrange, $x \geq \frac{a_1 + a_n}{2} \geq a_1$ and then $\max_i\{|a_i - x|\} \geq (x - a_1) \geq \frac{a_1 + a_n}{2} - a_1 = \frac{a_n - a_1}{2}$. When $x = \frac{a_1 + a_n}{2}$ (the midrange) each of these is equality so we get $\max_i\{|a_i - x|\} \geq \max_i |a_i - \frac{a_1 + a_n}{2}| = \frac{a_n - a_1}{2}$ and the midrange is a minimizer.

28. Given pairs of data points $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ consider approximating lines of the form $y = mx + b$.
(a) The error e_i for the i^{th} pair is the distance between y_i and the height (y value) of the line at x_i. Write this expression. (It will involve an absolute value.)
(b) Formulate a linear programming problem whose solution will give the best L_1 line. That is, the line that minimizes $\sum_{i=1}^{n} |e_i|$. You answer should not involve the e_i. It should be expressed in terms of the x_i, y_i and some variables that you introduce. State clearly which are variables and which are given values in your formulation and what the slope and intercept of the 'best' line

are in terms of an optimal solution to the problem you formulate.
(c) Formulate a linear programming problem whose solution will give the best L_∞ line. That is, the line that minimizes $\max|e_i|$. You answer should not involve the e_i. It should be expressed in terms of the x_i, y_i and some variables that you introduce. State clearly which are variables and which are given values in your formulation and what the slope and intercept of the 'best' line are in terms of an optimal solution to the problem you formulate.
(d) For both (b) and (c) write down the specific linear programs that you get for the data $(4,2),(13,\pi),(-3,-8)$.

(a) $|e_i| = |y_i - (mx_i + b)| = |y_i - mx_i - b|$.
(b) Use variables f_i for $i = 1,2,\ldots,n$ and m and b. To make sure that $f_i \geq |y_i - mx_i - b|$ we use inequalities $f_i \geq y_i - mx_i - b$ and $f_i \geq -(y_i - mx_i - b)$. If these are the only conditions and we minimize the sum of the f_i we will in fact have equality in an optimal solution and thus be minimizing the sum of the errors. So for the following linear programming problem in the variables $m, b, f_1, f_2, \ldots, f_n$ the values of m and b in an optimal are the slope and intercept of a best L_1 line. We rearrange the inequalities so that the variables are on the left side and given values on the right. (So the given x_i are coefficients for the variable m and the given y_i are the right hand sides.

$$\begin{array}{lrcrcrcrl}
\text{min} & 0m & + & 0b & + & \sum_{i=1}^n f_i & & & \\
\text{subject to} & x_i m & + & b & + & f_i & \geq & y_i & \text{for } i = 1,2,\ldots,n \\
 & -x_i m & - & b & + & f_i & \geq & -y_i & \text{for } i = 1,2,\ldots,n
\end{array}$$

(b) Use variables z and m and b. To make sure that $z \geq |y_i - mx_i - b|$ we use inequalities $z \geq y_i - mx_i - b$ and $z \geq -(y_i - mx_i - b)$. If these are the only conditions and we minimize z we will in fact have z equal to the maximum of the absolute values of the errors. So for the following linear programming problem in the variables m, b, z the values of m and b in an optimal are the slope and intercept of a best L_∞ line. We rearrange the inequalities so that the variables are on the left side and given values on the right. (So the given x_i are coefficients for the variable m and the given y_i are the right hand sides.

$$\begin{array}{lrcrcrcrl}
\text{min} & 0m & + & 0b & + & z & & & \\
\text{subject to} & x_i m & + & b & + & z & \geq & y_i & \text{for } i = 1,2,\ldots,n \\
 & -x_i m & - & b & + & z & \geq & -y_i & \text{for } i = 1,2,\ldots,n
\end{array}$$

(d) For the L_1 line we get

$$\begin{array}{lrcrcrcrcrcr}
\text{min} & 0m & + & 0b & + & f_1 & + & f_2 & + & f_3 & & \\
\text{subject to} & 4m & + & b & + & f_1 & & & & & \geq & 2 \\
 & 13m & + & b & & & + & f_2 & & & \geq & \pi \\
 & -3m & + & b & & & & & + & f_3 & \geq & -8 \\
 & -4m & - & b & + & f_1 & & & & & \geq & -2 \\
 & -13m & - & b & & & + & f_2 & & & \geq & -\pi \\
 & 3m & - & b & & & & & + & f_3 & \geq & 8
\end{array}$$

For the L_∞ line we get

$$
\begin{array}{rrcrcrcr}
\min & 0m & + & 0b & + & z & & \\
\text{subject to} & 4m & + & b & + & z & \geq & 2 \\
& 13m & + & b & + & z & \geq & \pi \\
& -3m & + & b & + & z & \geq & -8 \\
& -4m & - & b & + & z & \geq & -2 \\
& -13m & - & b & + & z & \geq & -\pi \\
& 3m & - & b & + & z & \geq & 8
\end{array}
$$