Exam 2 Review SOLUTIONS, Math 250

1. True or False, and explain:

   (a) In a nonlinear program, the domain D is convex.

   False- not necessarily. In the nonlinear case, we might have very complex curves and complex domains.

   (b) It is possible that a solution to a linear program is in the middle of an edge to the domain.

   False. We showed in class than a solution to a *linear* program will occur at a vertex along the feasible set. This was because the objective function, $\mathbf{c}^T\mathbf{x}$, restricted to any line segment, has its maximum and minimum at the endpoints of that line segment.

   (c) Let the domain D of a nonlinear program be given as $A\mathbf{x} \leq \mathbf{b}$. It is possible that a solution is in the middle of an edge of D.

   This is possible, and we saw it in one of the examples done in class.

   (d) The following set of points forms a convex set:

$$\left\{ \mathbf{x} : \begin{bmatrix} 3 & 4 & -1 \\ 0 & 2 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} \leq \begin{bmatrix} 5 \\ -1 \end{bmatrix} \right\}$$

   TRUE: Any (nonempty) set of the form $Ax \leq b$ forms a convex set.

2. Discussion/Short Answer:

   (a) State the "Fundamental Theorem of Linear Programming":

   Given $f(x) = C^T x$ such that $Ax \leq b$, which is nonempty, then there is a maximum and minimum value of f, and it is attained at a vertex of the polytope (or the feasible set).

   (b) State the Extreme Value Theorem: Let f be continuous over a closed and bounded domain D. Then f attains a maximum and minimum value on D.

   (c) What is a "linear program"? It is an optimization problem of the form: $\max_x c^T x$ such that $Ax \leq b$. (You could also use the min instead of the max).

   (d) Finish the definition: A set D is convex if: one takes any two points in D, all the points in between them are also in D. Algebraically, if $\mathbf{x}$ and $\mathbf{y}$ are in D, then so is $t\mathbf{x} + (1-t)\mathbf{y}$.

   (e) Under what conditions may we get no solution to a linear program? Does this change for a nonlinear program?

   It might be that the domain D forms an inconsistent set of equations (there is no feasible set). It might also be that D is unbounded. That's it, since if D is a closed and bounded set, then the linear program must have a solution.

   It doesn't change for a nonlinear program- the given constraint equations might be inconsistent, or the domain restrictions may form an unbounded set.

   (f) If $f : \mathbb{R}^n \to \mathbb{R}^n$, then the linearization of f at $\mathbf{x} = \mathbf{a}$ is:

$$L(x) = f(\mathbf{a}) + Jf(\mathbf{a})(\mathbf{x} - \mathbf{a})$$

   (g) If $f : \mathbb{R}^n \to \mathbb{R}$, then the linearization of f at $\mathbf{x} = \mathbf{a}$ is:

$$L(\mathbf{x}) = f(\mathbf{a}) + \nabla f(\mathbf{a}) \cdot (\mathbf{x} - \mathbf{a})$$

   (Note that the last term is a dot product).

Given the same f, the second order Taylor approximation is:

$$f(\mathbf{x}) \approx f(\mathbf{a}) + \nabla f(\mathbf{a}) \cdot (\mathbf{x}-\mathbf{a}) + \frac{1}{2}(\mathbf{x}-\mathbf{a})^T Hf(\mathbf{a})(\mathbf{x}-\mathbf{a})$$

(h) If we state that, "Given $g(x,y) = k$, it is possible under certain conditions to write this as $y = \hat{g}(x)$", what theorem are we using? The implicit function theorem.

(i) Let $\mathbf{x} \in \mathbb{R}^n$. Define the $p-$norm and $\infty-$norm for $\mathbf{x}$.

$$\|\mathbf{x}\|_p = (|x_1|^p + |x_2|^p + \ldots + |x_n|^p)^{1/p}$$
$$\|\mathbf{x}\|_\infty = \max_i |x_i|$$

(j) When should we use which error to compute our line of "best" fit?
From our experiments, we saw that:

- The $1-$norm was least sensitive to outliers.
- The $2-$norm works best if the error follows a "normal" distribution.
- The $\infty-$norm works best if the error is uniform.

(k) What is a neural network? What is it used for?
A neural network is generally a method for building functions from $\mathbb{R}^n$ to $\mathbb{R}^m$. It uses nonlinear optimization to find the best parameters.

3. Suppose our data set consists of the 3 points: $(1,2)$, $(2,1)$, $(3,4)$. We want to find the line of best fit.

(a) Write (explicitly, without $\sum$) the error function using the $2-$norm and find the line of best fit (by hand).
$E(m,b) = (2-m-b)^2 + (1-2m-b)^2 + (4-3m-b)^2$ To find the line of best fit, differentiate, set the derivatives to zero and solve for m, b:

$$2(2-m-b)(-1) + 2(1-2m-b)(-2) + 2(4-3m-b)(-3) = 0 \Rightarrow 14m + 6b = 16$$

$$2(2-m-b)(-1) + 2(1-2m-b)(-1) + 2(4-3m-b)(-1) = 0 \Rightarrow 6m + 3b = 7$$

From these two equations, we get $m = 1$, $b = 1/3$.

(b) Write (explicitly, without $\sum$) the error function using the $1-$norm. Translate the problem into a linear programming problem (without explicit use of the ϵ_i, although you may use them to get to your final answer).

$$E(m,b) = |2-m-b| + |1-2m-b| + |4-3m-b|$$

Its always best to consider what will be your basic set of variables (so you can construct $\mathbf{c}^T\mathbf{x}$). Here, we will have m, b, and a_1, a_2, a_3, which will be our absolute values. Given these variables, then:

$$\mathbf{c} = [0,0,1,1,1]^T$$

where the constraints give the relationships between the variables: $a_i \geq \epsilon_i$ and $a_i \geq -\epsilon_i$:

$$a_1 \geq 2-m-b \quad \text{and} \quad a_1 \geq -2+m+b$$

$$a_2 \geq 1-2m-b \quad \text{and} \quad a_2 \geq -1+2m+b$$

$$a_3 \geq 4-3m-b \quad \text{and} \quad a_3 \geq -4+3m+b$$

so $A\mathbf{x} \leq \mathbf{b}$ is given by:

$$\begin{bmatrix} -1 & -1 & -1 & 0 & 0 \\ -2 & -1 & 0 & -1 & 0 \\ -3 & -1 & 0 & 0 & -1 \\ 1 & 1 & -1 & 0 & 0 \\ 2 & 1 & 0 & -1 & 0 \\ 3 & 1 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} m \\ b \\ a_1 \\ a_2 \\ a_3 \end{bmatrix} \leq \begin{bmatrix} -2 \\ -1 \\ -4 \\ 2 \\ 1 \\ 4 \end{bmatrix}$$

(c) Write (explicitly, without $\sum$) the error function using the $\infty-$norm. Translate the problem into a linear programming problem.

In this case,

$$E(m,b) = \max\ \{|2-m-b|, |1-2m-b|, |4-3m-b|\}$$

Now, let ϵ_M be the max of these three quantities. Our basic set of variables is now m, b, ϵ_M, and our vector $\mathbf{c} = [0,0,1]^T$.

For the relationships: $-\epsilon_M \leq y_i - mx_i - b \leq \epsilon_M$, so for each data point we have:

$$mx_i + b - \epsilon_M \leq y_i \qquad \text{and} \qquad -mx_i - b - \epsilon_M \leq y_i$$

so $A\mathbf{x} \leq \mathbf{b}$ is given by:

$$\begin{bmatrix} -1 & -1 & -1 \\ -2 & -1 & -1 \\ -3 & -1 & -1 \\ 1 & 1 & -1 \\ 2 & 1 & -1 \\ 3 & 1 & -1 \end{bmatrix} \begin{bmatrix} m \\ b \\ \epsilon_M \end{bmatrix} \leq \begin{bmatrix} -2 \\ -1 \\ -4 \\ 2 \\ 1 \\ 4 \end{bmatrix}$$

4. Solve the linear program graphically: $\max_{x,y}\ x+y$ so that $x \geq 0$, $y \geq 0$, $3x+2y \leq 5$, $2x+3y \leq 5$.

   If the objective function is replaced by $\alpha x + y$ with $\alpha > 0$, for what values of α is the solution still optimal?

   Your graph should be bounded by the vertices $(0,0)$, $(5/3,0)$, $(1,1)$, $(0,5/3)$. We get the maximum then of 2 at $(1,1)$.

   Note that the current slope is -1, and the slopes of the other two sides are $-3/2$ and $-2/3$. If α stays greater than $-3/2$, we retain the same maximizer, but at $-3/2$, we get an infinite number of solutions (all points on the upper boundary). If α drops below that, then $(0,5/3)$ becomes the new maximizer.

   Similarly, if α stays below $-2/3$, then we get the same maximizer. At $-2/3$, we get a line of maximizers (the line to the left), and if α increases beyond that, $(5/3,0)$ becomes the new maximizer.

   In summary, $(1,1)$ is the maximizer as long as $-\frac{3}{2} < \alpha < -\frac{2}{3}$.

5. Consider the linear program: $\max 3x+y$ such that $x,y \geq 0$, $x+y \geq 2$, $x-y \leq 2$ and $y \leq 3$.

   (a) Graph the feasible region and label all vertices of the polytope. Check your answer on Maple.
   (b) Solve the LP graphically. See the solution below.
   (c) Show that this is the solution by using the vertices.
   The solution is: $x = 5, y = 3$

6. Use Lagrange multipliers to find the maximum of $4x+6y$ if $x^2+y^2 = 16$.

   We have the system of equations:

$$\begin{aligned} 4 &= \lambda 2x \\ 6 &= \lambda 2y \\ x^2+y^2 &= 16 \end{aligned}$$

We can multiply the first equation by y and the second equation by x to get that $y = \frac{3}{2}x$. Substitute this into the third equation to solve for y.

Another way to get the same solutions: Assuming that $\lambda \neq 0$, we get that $x = \frac{2}{\lambda}$, $y = \frac{3}{\lambda}$. Substitute these into the third equation and solve for λ.

Using these, we get

$$x = \pm\frac{8}{\sqrt{13}}, \quad y = \pm\frac{12}{\sqrt{13}}$$

so the maximum will be found by taking the positive solutions for x, y: $104/\sqrt{13}$.

7. Find the critical points of $f(x, y) = 3x - x^3 - 2y^2 + y^4$.

   We check for where the gradient is zero. In this example,

$$\nabla f = [3 - 3x^2,\ -4y + 4y^3]$$

   Altogether, we have six points:

$$(1, -1), (1, 0), (1, 1), (-1, -1), (-1, 0), (-1, 1)$$

8. Find the global maximum of $f(x, y) = 1 + xy - x - y$ on the domain D bounded by the parabola $y = x^2$ and the line $y = 4$.

   We first check the critical points of f:

$$\nabla f = [y - 1, x - 1]$$

   so that the only critical point is $(1, 1)$, and $f(1, 1) = 0$.

   We now check the boundary in two parts:

   - $y = 4$, $-2 \leq x \leq 2$. Now $f(x) = 1 + 4x - x - 4 = -3 + 3x$. On the interval $[-2, 2]$, this function has its maximum at $x = 2$. For this point, $f(2, 4) = -3 + 6 = 3$.
   - $y = x^2$, $-2 \leq x \leq 2$. Now $f(x) = 1 + x^3 - x - x^2$, or $f(x) = x^3 - x^2 - x + 1$. To find its maximum, check the critical points and the endpoints:

$$f'(x) = 3x^2 - 2x - 1 \Rightarrow x = 1, -1/3$$

     Now we check the value of f at $x = -2, -1/3, 1, 2$:

$$f(-2) = -9, f(-1/3) = 32/27 \approx 1.19, f(1) = 0, f(2) = 3$$

     so once again the maximum is at $x = 2, y = 4$.

   Overall, we have found the maximum value to be 3, which occurs at $(2, 4)$.

9. Recall that for a 2×2 matrix, we can compute the inverse directly:

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \quad A^{-1} = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

   Use this formula to compute one step of multidimensional Newton's method, if $f(x, y)$ is given below, and the inital guess is $(1, 1)$.

$$f(x, y) = \begin{bmatrix} x^2 + 2y \\ 3xy - y^2 \end{bmatrix}$$

In this case, we use the formula:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - (Jf(\mathbf{x}_k))^{-1} f(\mathbf{x}_k)$$

The Jacobian matrix is given by:

$$Jf(x,y) = \begin{bmatrix} 2x & 2 \\ 3y & 3x-2y \end{bmatrix}, \quad Jf(1,1) = \begin{bmatrix} 2 & 2 \\ 3 & 1 \end{bmatrix} \text{ and } f(1,1) = [3,2]^T$$

so that at the point $(1,1)$,

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \frac{1}{4} \begin{bmatrix} 1 & -2 \\ -3 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 3/4 \\ -1/4 \end{bmatrix}$$

10. Let $f(x,y,z) = \dfrac{x}{y} + \dfrac{y}{z}$. At the point $(4,2,1)$, find the maximum rate of change of f and the direction in which it occurs.

    The maximum rate of change of f occurs in the direction of the gradient. The rate at which the change occurs is the size (or norm) of the gradient:

    $$\nabla f = \left[\frac{1}{y},\ -\frac{x}{y^2} + \frac{1}{z},\ -\frac{y}{z^2}\right]$$

    At the point $(4,2,1)$, we have that this direction is $\left[\frac{1}{2}, 0, -2\right]$ and the rate of change is $\frac{\sqrt{17}}{2}$

11. Formulate the following problem as an LP. Do not solve it:

    A confectioner manufactures two kinds of candy bars; "ProteinPlus", that has no carbohydrates, and "SugarPlus", with no fat. ProteinPlus sells for a profit of 40 cents per bar, and SugarPlus sells for a profit of 50 cents per bar. The candy is processed in three main operations; blending, cooking and packaging. The following table records the average time in minutes required by each bar for each of the processing operations:

	Blending	Cooking	Packaging
ProteinPlus	1	5	3
SugarPlus	2	4	1

    During each production run, the blending equipment is available for a maximum of 12 machine hours, the cooking equipment is available for at most 30 machine hours, and the packaging equipment for no more than 15 hours. If this machine time can be allocated to the making of either type of candy at all times that it is available, determine how many boxes of each kind the confectioner should make in order to realize the maximum profit. See the solution below.

12. (Referring to the previous problem) Solve the LP problem. The solution is given by $P = 120, S = 300$. You can check this and plot the feasible region in Maple:

```
restart;
with(simplex):
CS:=P+2*S<=12*60, 5*P+4*S<=30*60, 3*P+S<=15*60;
maximize(0.4*P+0.5*S,[CS],NONNEGATIVE);
with(plots):
inequal({CS},P=0..800,S=0..800,optionsfeasible=(color=
white), optionsexcluded=(color=gray));
```

13. Suppose we are given a neural network with 2 input nodes, 3 nodes in the hidden layer, and 1 output node.

Typo: See W^2 below (it was missing a value), and the input should be $(1,1)$

(a) Our domain is in $\mathbb{R}^2$ Our range is in $\mathbb{R}$.

(b) How many weights and biases are there altogether? 13 (See below)

(c) For the input and output layers, assume $\sigma(x) = x$, and for the hidden layer, assume $\sigma(x) = \arctan(x)$. If the weights and biases are:

$$W^{(1)} = \begin{bmatrix} -1 & 0 \\ 0 & 1 \\ 1 & -1 \end{bmatrix} \qquad \mathbf{b}^{(1)} = \begin{bmatrix} 0.5 \\ 0.1 \\ -1 \end{bmatrix} \qquad W^{(2)} = \begin{bmatrix} 2 & -1 & 1 \end{bmatrix} \qquad \mathbf{b}^{(2)} = 1$$

compute the prestates and states of each node if the data point $(1,1)$ is input to the network.
The prestates and states of the input layer are the data values themselves, $1,1$
The hidden layer prestates are:

$$-0.5, 1.1, -1.0$$

The hidden layer states are:

$$-0.464, 0.833, -0.785$$

The output layer has prestate and state: $-2.546 + 1 = -1.546$.

14. Fill in the blank with *perpendicular* or *parallel*. Also answer the associated questions.

(a) Let $f : \mathbb{R} \to \mathbb{R}^2$. The line at $f(2)$ in the direction of $f'(2)$ is parallel (or tangent) to the curve $f(t)$.

(b) Let $z = f(x,y)$. The gradient of f at (a,b) is perpendicular to a level curve of f.
Show your answer is correct, if $z = f(x,y) = x^2 + 3xy$ at the point $(1,1)$. Which level curve are you on?
We are on the level curve $f(x,y) = f(1,1) = 1+3 = 4$. In this particular example, this would mean that we check to see if the gradient, $[2x+3, 3x]$ at $(1,1)$ giving a gradient of $[5,3]$ is perpendicular to the surface $x^2 + 3xy = 4$ at $(1,1)$. To check this, we need the slope of the tangent line to the curve at $(1,1)$ which we can get by implicit differentiation:

$$2x + 3y + 3xy' = 0 \Rightarrow y' = -\frac{2x+3y}{3x}$$

At $(1,1)$, the slope is $-5/3$. A line in the direction of the gradient would have a slope of $3/5$, so they are indeed perpendicular.

(c) Let $k = f(x,y)$ be a level curve that can be written as $y = g(x)$. Then at a point (a,b), the gradient of f at (a,b) is *perpendicular* to the line $y - b = g'(a)(x-a)$.

(d) Suppose $z = f(x,y)$ with the domain restriction that $g(x,y) = k$. If the optimum of f (with the restricted domain) is (a,b), then:

- Does $\nabla f(a,b) = 0$? Not necessarily, although it might happen by chance. Generally, we do not expect f to actually have critical points on the set $g(x,y) = k$.
- $\nabla f(a,b)$ is *parallel* to $\nabla g(a,b)$

15. Compute the gradient and Hessian of $f(x,y) = 5 - 7y + 8x^2y + 3y^2$.

$$\nabla f = [16xy \quad -7 + 8x^2 + 6y]$$

$$Hf = \begin{bmatrix} 16y & 16x \\ 16x & 6 \end{bmatrix}$$

16. Write the equations necessary to solve the following problem (but do not solve it):

    Suppose there are n points on the unit sphere. Find the positions of the points so that they all have maximal separation from each other. (This is from a physics problem, where the points are electrons, and they repel each other).

    One way to write this problem: Let $\{(x_i, y_i, z_i)\}_{i=1}^{n}$ be the unknown points on the sphere. Then:

$$\min \sum_{i=1}^{n} \sum_{j=i+1}^{n} (x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2$$

    such that:

$$x_i^2 + y_i^2 + z_i^2 = 1, \quad i = 1 \ldots n$$

    You might notice that this does not have a unique solution, as there are an infinite number of ways of placing the n points on a sphere. One way to approach the solution of this problem would be to initialize the n points at random, then update the positions until we get to an approximate solution.

    This was a problem I actually ran into when I was working on a method to "triangulate" the surface of a lung.