Math 350 Exam 2 Review Questions

1. What is a Voronoi diagram?

2. Is data clustering an example of supervised or unsupervised learning? Explain (and give an explanation of the overall problem).

3. How is the rank computed when we construct either the reduced SVD or the pseudoinverse?

4. Given the function $f(x, y)$, show that the direction in which f decreases the fastest from a point (a, b) is given by the negative gradient (evaluated at (a, b)). Hint: A certain dot product can be related to the cosine of the angle between the vectors.

5. Illustrate the technique of gradient descent using

$$f(x, y) = x^2 + y^2 - xy + 2$$

   (a) Find the minimum.

   (b) Use the initial point $(1, 0)$ and $\alpha = 0.1$ to perform one step of gradient descent (use your calculator).

   (c) Same problem, but use line search to find the optimal step size (start at $(1, 0)$ again).

   (d) Calculate one step of multivariate Newton's method on the gradient (at the point $(1, 0)$) using a step size of $\alpha = 1/3$.

6. If

$$f(t) = \begin{bmatrix} 3t - 1 \\ t^2 \end{bmatrix}$$

   find the tangent line to f at $t = 1$.

7. If $f(x, y) = x^2 + y^2 - 3xy + 2$, find the linearization of f at $(1, 0)$.

8. How did we define the notion of "best" in the best basis? To help, suppose we have an arbitrary orthonormal basis $\{\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_n\}$ and data $\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_p\}$.

9. If C is the covariance matrix given below, find the maximum and minimum of $F(\phi)$, and give the ϕ for which the maximum occurs (we may assume ϕ is not the zero vector, and that ϕ is a vector with 2 elements).

$$F(\phi) = \frac{\phi^T C \phi}{\phi^T \phi} \qquad \text{for } C = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$$

   (Hint: You may find it easily using our theorems)

10. Given data in $\mathbb{R}$: $x_1, \ldots, x_p$, show that, if we define the function E below:

$$E(m) = \frac{1}{p} \sum_{i=1}^{p} (x_i - m)^2$$

   then find the value of m that minimizes E.

11. Give the algorithm for $k-$means clustering:

12. Give the cluster update rule for Kohonen's self organizing map.

13. Give the cluster update rule for Neural Gas.

14. What is the main difference between SOM and Neural Gas?

15. Here is one data point. There are three centers in the matrix C which have a linear topology- That is, I gives the one-dimensional representation of each cluster center.

    Perform one update of the centers using Kohonen's SOM update rule, assuming that $\epsilon = \lambda = 1$ (unrealistic, but easier to do by hand):

$$\mathbf{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \qquad C = \begin{bmatrix} -1 & 1 & 2 \\ 1 & 0 & 3 \end{bmatrix} \qquad I = [1, 6, 3]$$

    Also, for the distance in the plane, use the "taxicab" or "Manhattan" metric:

$$d(\mathbf{a}, \mathbf{b}) = |a_1 - b_1| + |a_2 - b_2|$$

16. Same as the previous problem, but update using the Neural Gas algorithm (assume all the centers are connected and ignore the age). Use $\epsilon = \lambda = 1$ (unrealistic, but this is by hand). For the metric in the plane, again use the taxicab metric.
17. In the DBSCAN algorithm, points are classified into three different groups- What were the groups, and how were the groups defined?
18. What are the training parameters that must be set before using the DBSCAN algorithm?
19. Define what it means for q to be (directly) density-reachable from p (in the DBSCAN context).
20. What are the "inputs" to Kohonen's SOM? (That is, what information needs to be provided to the algorithm)?
21. Similarly, what are the inputs to Neural Gas?
22. Let $f(x, y) = 3x^2 + xy - y^2 + 3x - 5y$.
    (a) From the point $(1, 1)$, in which direction is f increasing the fastest?
    (b) Find the critical point of f.
    (c) Compute the Hessian of f and determine if f has a local max or min at the critical point (recall that we compute eigenvalues, but we only need the signs of the eigenvalues).
23. Let $f(x, y) = 3xy + x^2$.
    (a) Linearize f about the point $(1, 1)$.
    (b) Compute the Hessian of f.
    (c) Show that Newton's Method, starting at $(1, 1)$, converges in one step.