Review Questions

1. (Calculator, or by hand) Use two steps of the bisection algorithm on $f(x) = x^2 - 2$ on the interval $[0, 2]$. Be sure you follow the steps.

2. (Calculator, or by hand) Use two steps of Newton's Method on $f(x) = x^2 - 2$ with $x_0 = 1$.

   (On the exam, the numbers will work out without a calculator)

3. Define a "voronoi cell" and its relation to data clustering.

4. Explain the roles that ϵ and λ play in the Neural Gas algorithm.

5. Let x_i be a set of p real numbers. Prove that the value of c that minimizes the following quantity:

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

   is the mean.

6. Here are 5 points in the matrix X. Initialize two centers as the first two columns of X, compute the initial distortion error, then perform 1 update using $k-$means to get the number cluster centers.

$$X = \begin{bmatrix} -1 & 1 & 1 & -2 & -1 \\ 1 & 0 & 2 & 1 & -1 \end{bmatrix}$$

7. Given the data vector $\mathbf{x}$ below and the three centers in C, update the set of centers using Neural Gas, with $\epsilon = \lambda = 1$ (not realistic, but since we're doing it by hand, we'll use easy numbers).

$$\mathbf{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \qquad C = \begin{bmatrix} -1 & 1 & 2 \\ 1 & 0 & 3 \end{bmatrix}$$

8. Given a linear network that maps $\mathbb{R}^2 \to \mathbb{R}$,

$$y = w_1 x_1 + w_2 x_2 + b = f(x_1, x_2)$$

   and given the error function using data point $(\mathbf{x}, t)$

$$E(w_1, w_2, b) = (t - y)^2$$

   find an expression for the partial derivatives of E with respect to each parameter.

9. Given the function

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

   use the second derivatives test to classify the critical points (you may alternatively use the eigenvalues of the Hessian).

10. Suppose we are given the function

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

    We want to find a critical point of f using the multivariate Newton's method. Just perform one step, with initial point $(1, 1)$ and step size $\alpha = 0.1 = 1/10$.

11. (From exercise 1 on p 116 of the optimization notes)

    Below you are given four data points. We want to find the line of best fit, $y = mx + b$.

(a) Write out the error function that we will minimize.

(b) Write the gradient of the error function.

(c) With an initial guess of $m = 0$, $b = -1$, and a learning rate (or step size) of $0.1 = 1/10$, perform one step of gradient descent.

(d) Using the third point $(2, 3)$, perform one step of stochastic gradient descent using $(1, 1)$ and learning rate 0.1, as before.

(A bit of computation here- The main point is to understand the steps involved in the algorithms.)

12. How do you change an affine equation into a linear equation? That is, change the matrix-vector equation:

$$A\mathbf{x} + \mathbf{b} = \mathbf{y}$$

into an equivalent linear equation, $\hat{A}\hat{\mathbf{x}} = \mathbf{y}$:

13. What is the Widrow-Hoff update rule? You may write it either in matrix form or in scalar form.

14. In pattern classification, suppose I have data in the plane that I want to divide into 5 classes. Would I want to build a pattern classification function f so that the range is the following set:

$$\{1, 2, 3, 4, 5\}$$

Why or why not? If not, what might be a better range?

15. 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)).

16. If

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

find the tangent line to f at $t = 1$ (this linearizes f at $t = 1$).

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

18. Give the algorithm for $k-$means clustering.

19. Give the algorithm for Neural Gas.

20. In DBSCAN, we classify each data point into three sets. What are they?

21. In DBSCAN, what are the two main parameters that must be set?

22. For a linear neural net, given just one data point:

$$X = \begin{bmatrix} 2 \\ -1 \end{bmatrix} \qquad T = [1]$$

initialize W and $\mathbf{b}$ as an appropriately sized arrays of ones, then perform one update using the Widrow-Hoff rule with step size $\alpha = 0.1$.