Review Questions (Exam 2)

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

2. Define the "confusion matrix", and how it is used.

3. What is the basic update rule we use for all our parameters? Hint: We want to go from α_{initial} to α_{final} in some number (MaxIters) of steps.

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

5. Show that, for all numbers μ, the value that minimizes the (squared) distortion error for a single cluster is the (arithmetic) mean. You may assume your data is one dimensional, and that you have only one cluster.

6. Here are 5 points in the matrix X. Initialize the two centers as the first two columns of X, then perform 1 update, and show there is a decrease in the distortion error.

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

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. Show that, for the line of best fit, the normal equations produce the same equations as minimizing an appropriate error function. To be more specific, set the data as $(x_1, t_1), \ldots, (x_p, t_p)$ and define the error function first. Minimize the error function to find the system of equations in m, b. Show this system is the same you get using the normal equations.

9. Given data:

x	-1	0	1
y	2	1	1

   (a) Give the matrix equation for the *line of best fit.*

   (b) Compute the normal equations.

   (c) Solve the normal equations for the slope and intercept.

10. Use the data in Exercise (9) to find the parabola of best fit: $y = ax^2 + bx + c$. (NOTE: Will you only get a least squares solution, or an actual solution to the appropriate matrix equation?)

11. What is Hebb's rule (the biological version- you can paraphrase)?

12. What is the Widrow-Hoff learning rule? How is it related to Hebb's rule?

13. Let $W = \begin{bmatrix} 1 & 0 & 1 \\ -1 & 1 & 0 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. If $\mathbf{x} = [-1, 0, 1]^T$ and $\mathbf{t} = [2, 3]^T$, use Widrow-Hoff to update $W, \mathbf{b}$ one time using a learning rate of 1 (This is too big of a learning rate to actually use, but it will make your computations easier).

14. Let $\mathbf{x} = [1, 2, 1]^T$. Find the matrix $\mathbf{x}\mathbf{x}^T$, its eigenvalues, and eigenvectors. (Also, think about what happens in the general case, where a matrix is defined by $\mathbf{x}\mathbf{x}^T$).

15. Suppose $\mathbf{x}$ is a vector containing n real numbers, and we understand that $m\mathbf{x} + b$ is Matlab-style notation (so we can add a vector to a scalar, done component-wise).

    (a) Find the mean of $\mathbf{y} = m\mathbf{x} + b$ in terms of the mean of $\mathbf{x}$.

    (b) Show that, for fixed constants a, b, $\text{Cov}(\mathbf{x} + a, \mathbf{y} + b) = \text{Cov}(\mathbf{x}, \mathbf{y})$

    (c) If $\mathbf{y} = m\mathbf{x} + b$, then find the covariance and correlation coefficient between $\mathbf{x}$ and $\mathbf{y}$.

16. Show that the affine mapping: $f(\mathbf{x}) = W\mathbf{x} + \mathbf{b}$ can be written as a linear mapping $\hat{W}\hat{\mathbf{x}}$ for an appropriate $\hat{W}$ and $\hat{\mathbf{x}}$

17. What does "training" mean in terms of our mathematical model?

18. If we use all the data we have at once, what kind of training are we doing? If we learn one data point at a time, what kind of training are we doing?

19. Suppose I have some data in $\mathbb{R}^3$ that belongs to 4 different classes. Do I want my targets to be the real numbers $1, 2, 3, 4$, or are there better ways to build the target values?

20. Given the function $z = f(x, y)$, 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)).

21. Illustrate the technique of gradient descent using

$$f(x, y) = x^2 + y^2 - 3xy + 2$$

    (a) Find the minimum.

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

22. Suppose we have a subspace W spanned by an orthonormal set of non-zero vectors, $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$, each is in $\mathbb{R}^{1000}$. If a vector $\mathbf{x}$ is in W, then there is a low dimensional (three dimensional in fact) representation of $\mathbf{x}$. What is it?

23. Let the matrix A be defined below.

$$A = \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{bmatrix}$$

   (a) Find the psuedoinverse of A

   (b) Using the A from the previous exercise, consider the vector $[-1, 0, 1]^T$. Is the vector in the column space of A? If so, provide its coordinates with respect to the columns of A (for the basis).

   (c) What happens if we try to project $[1, -2, 1]^T$ into the column space of A? Explain in terms of fundamental subspaces.

24. Consider the underdetermined "system of equations": $x+3y+4z = 1$. In matrix-vector form $A\mathbf{x} = \mathbf{b}$, write the matrix A first.

   (a) What is the dimension of each of the four fundamental subspaces?

   (b) Find bases for the four fundamental subspaces.

   (c) Find a solution with at least 2 zeros (the slash command in Matlab looks for answers with the most zeros).

   (d) Find a 3×3 matrix P so that given a vector $\mathbf{x}$, $P\mathbf{x}$ is the projection of $\mathbf{x}$ into the row space of A.

25. (Eigenvalues) Find the eigenvalues and eigenvectors for $A = \begin{bmatrix} 0 & 1 \\ 2 & 1 \end{bmatrix}$

26. (Eigenvalues) Verify the 4 results of the Spectral Theorem for $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$

27. (Eigenvalues) If λ_i is an eigenvalue of $A^T A$, then show that $\lambda_i \geq 0$ by showing that, if $\mathbf{v}_i$ is an eigenvector for λ_i, then $\|A\mathbf{v}_i\|^2 = \lambda_i$ (and lengths cannot be negative).

28. (Eigenvalues) If $\mathbf{v}_i$ and $\mathbf{v}_j$ are eigenvectors corresponding to distinct eigenvalues of $A^T A$, then show that $A\mathbf{v}_i \perp A\mathbf{v}_j$.

29. (Eigenvalues) Suppose that $\lambda_i, \mathbf{v}_i$ are eigenvalue/eigenvectors for a symmetric matrix $A^T A$ (so the Spectral Theorem applies). Prove that, if $\boldsymbol{x} = \alpha_1 \boldsymbol{v}_1 + \ldots \alpha_n \boldsymbol{v}_n$, then

$$\|A\boldsymbol{x}\|^2 = \alpha_1^2 \lambda_1 + \ldots + \alpha_n^2 \lambda_n$$

30. (Eigenvalues) Prove that if λ_i is an eigenvalue of $A^T A$, then λ_i is also an eigenvalue of AA^T (Hint: Let $\mathbf{u}_i = A\mathbf{v}_i$, where $\mathbf{v}_i$ is an eigenvector associated with λ_i).

31. (SVD) Compute the SVD by hand of the following matrices. Verify by computing $\sum \sigma_i \mathbf{u}_i \mathbf{v}_i^T$

$$\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \qquad \begin{pmatrix} 0 & 2 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$$

32. (SVD) Given that the SVD of a matrix was given in Matlab as:

```
>> [U,S,V]=svd(A)
U =
   -0.4346   -0.3010    0.7745    0.3326   -0.1000
   -0.1933   -0.3934    0.1103   -0.8886   -0.0777
    0.5484    0.5071    0.6045   -0.2605   -0.0944
    0.6715   -0.6841    0.0061    0.1770   -0.2231
    0.1488   -0.1720    0.1502   -0.0217    0.9619
S =
    5.72         0           0
         0       2.89        0
         0         0         0
         0         0         0
         0         0         0
V =
    0.2321   -0.9483    0.2166
   -0.2770    0.1490    0.9493
    0.9324    0.2803    0.2281
```

(a) Which columns form a basis for the null space of A? For the column space of A? For the row space of A?

(b) How do we "normalize" the singular values? In this case, what are they (numerically)?

(c) What is the rank of A?

(d) How would you compute the pseudo-inverse of A (do not actually do it):

(e) Let B be formed using the first two columns of U. Would the matrix $B^T B$ have any special meaning? Would BB^T?