Review Questions, Exam 1

1. What was the $N-$armed bandit problem? In particular, what were the two competing goals, and why were they "competing"?

2. In the $N-$armed bandit problem, how were the estimates of the payoffs, $Q_t(a)$, calculated?

3. There were four "strategies" that we implemented as algorithms to solve the $N-$armed bandit problem. What were they? Be sure to give formulas where appropriate.

4. Describe (in words) the greedy algorithm and the $\epsilon-$ greedy algorithm. Which is probably a better strategy?

5. Describe in words the softmax strategy. Be sure to include appropriate formulas, and describe what the parameter τ does.

6. What was the pursuit strategy (or "Win-Stay, Lose-Shift") for the $N-$armed bandit? Again, include appropriate formulas and describe what β does.

7. Matlab Questions:

   (a) What's the difference between a script file and a function?

   (b) What does the following code fragment produce?

   ```
   Q=[1 3 2 1 3];
   idx=find(Q==max(Q));
   ```

   (c) What is the difference between `x=rand;` and `x=randn;`

   (d) What will P be:

   ```
   x=[0.3, 0.1, 0.2, 0.4];
   P=cumsum(x);
   ```

   (e) What is the Matlab code that will:

      i. Plot $x^2 - 3x$ using 500 points, for $x \in [-1, 4]$

      ii. Compute the variance of data in a vector $\boldsymbol{x}$ (possibly varying in length). You can't use `var`!

      iii. Compute the covariance of data in a vector $\boldsymbol{x}$, and $\boldsymbol{y}$ of the same, but possibly varying length. You can't use `cov`!

8. Give mathematical formulas for the sample mean and sample variance. Give the formulas for covariance and correlation.

9. What is the definition of the covariance matrix to X (say that X has p vectors in $\mathbb{R}^n$, and X is $n \times p$). You can define it by saying what the (i, j)th term of the covariance matrix represents.

10. Find the orthogonal projection of the vector $\boldsymbol{x} = [1, 0, 2]^T$ to the plane defined by:

$$G = \left\{ \alpha_1 \begin{bmatrix} 1 \\ 3 \\ -2 \end{bmatrix} + \alpha_2 \begin{bmatrix} 3 \\ -1 \\ 0 \end{bmatrix} \text{ such that } \alpha_1, \alpha_2 \in \mathbb{R} \right\}$$

Determine the distance from $\boldsymbol{x}$ to the plane G.

11. If $[\boldsymbol{x}]_{\mathcal{B}} = (3, -1)^T$, and $\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix} \right\}$, what was $\boldsymbol{x}$ (in the standard basis)?

12. If $\boldsymbol{x} = (3, -1)^T$, and $\mathcal{B} = \left\{ \begin{bmatrix} 6 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -2 \end{bmatrix} \right\}$, what is $[\boldsymbol{x}]_{\mathcal{B}}$?

13. Let $\boldsymbol{a} = [1, 3]^T$. Find a square matrix A so that $A\boldsymbol{x}$ is the orthogonal projection of $\boldsymbol{x}$ onto the span of $\boldsymbol{a}$.

14. Determine the projection matrix that takes a vector $\mathbf{x}$ and outputs the projection of $\mathbf{x}$ onto the plane whose normal vector is $[1, 1, 1]^T$.

15. Find (by hand) the eigenvectors and eigenvalues of the matrix A:

$$A = \begin{bmatrix} 5 & -1 \\ 3 & 1 \end{bmatrix}, \qquad A = \begin{bmatrix} -2 & 1 \\ 1 & -2 \end{bmatrix}$$

16. Compute the orthogonal projector to the span of $\mathbf{x}$, if $\mathbf{x} = [1, 1, 1]^T$.

17. Let

$$U = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \\ 0 & 0 \end{bmatrix}, \qquad \mathbf{x} = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}$$

Find $[\mathbf{x}]_U$. Find the projection of $\mathbf{x}$ into the subspace spanned by the columns of U. Find the distance between $\mathbf{x}$ and the subspace spanned by the columns of U.

18. Show that $\text{Null}(A) \perp \text{Row}(A)$.

19. Show that, if X is invertible, then $X^{-1}AX$ and A have the same eigenvalues.

20. How do we "double-center" a matrix of data?

21. True or False, and give a short reason:

    (a) If the rank of A is 3, the dimension of the row space is 3.

    (b) If the correlation coefficient between two sets of data is 1, then the data sets are the same.

(c) If the correlation coefficient between two sets of data is 0, then there is no functional relationship between the two sets of data.

(d) If U is a 4×2 matrix, then $U^T U = I$.

(e) If U is a 4×2 matrix, then $UU^T = I$.

(f) If A is not invertible, then $\lambda = 0$ is an eigenvalue of A.

22. 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.

23. 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.

24. Use the data in Exercise (23) 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?)

25. 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, $\mathrm{Cov}(\mathbf{x} + a, \mathbf{y} + b) = \mathrm{Cov}(\mathbf{x}, \mathbf{y})$

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

26. 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?

27. 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.

28. Suppose $Q = [-0.5, 0, 0.5, 1.0]$. Use the softmax selection technique with $\tau = 0.1$ to compute the probabilities.

29. If $Q_1 < Q_2 < Q_3 < Q_4$ for 4 machines, how do the probabilities change (under softmax) as $\tau \to 0$? As $\tau \to 1$?

30. What is the win-stay, lose-shift (or pursuit) strategy? What are the update rules?

31. Suppose we play with three machines, and machine 3 is chosen and gives a big payout (enough to make $Q_t(3)$ the maximum). Update the probabilities for win-stay, lose-shift, if they are: $P_1 = 0.3, P_2 = 0.5, P_3 = 0.2$ and $\beta = 0.3$.

32. Suppose we have a genetic algorithm with 4 chromosomes, and current fitness values $[-1, 1, 2, 3]$. We want to construct the probability of choosing these chromosomes for mating. Calculate the probabilities (if possible) using the current ordering, if we use:

    (a) Normalized fitness values:

    (b) Rank order, normalized:

    (c) Softmax with $\tau = 1$.