SOLUTIONS: Sample Questions

1. Compute the SVD of the matrix $A = \begin{bmatrix} 2 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$

   SOLUTION: Generally to compute the SVD by hand we need the eigenvalues and eigenvectors of A^TA and AA^T. First, we'll look at A^TA:

$$A^TA = \begin{bmatrix} 4 & 0 \\ 0 & 0 \end{bmatrix}$$

   The eigenvalues are $4, 0$ and (since the matrix is already diagonalized), the eigenvectors are the columns of the identity matrix. Similarly,

$$AA^T = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

   The eigenvalues are $4, 0, 0$, and the eigenvectors are the columns of the identity matrix. Therefore, the singular values (for Σ) are $\sqrt{4} = 2$ and 0:

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

   For extra practice, what is the **reduced** SVD? What is the pseudo-inverse of A? (Answers at the end)

2. 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?

SOLUTION: The matrix A is 5×3, so the null space and row space are subspaces of $\mathbb{R}^3$. The column space (and null space of A^T) are in $\mathbb{R}^5$. The dimension of the column space (which is also the dimension of the row space) is 2. Now we can answer the questions:

- The last column of V is a basis for the 1-dimensional null space.
- The column space is spanned by the first two columns of U.
- The row space is spanned by the first two columns of V.

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

SOLUTION: Since $\sigma_1 = 5.72$ and $\sigma_2 = 2.89$, and $\sigma_1 + \sigma_2 = 8.61$, and

$$\hat{\sigma}_1 = \frac{\sigma_1}{\sigma_1 + \sigma_2} = 0.6643 \qquad \hat{\sigma}_1 = \frac{\sigma_2}{\sigma_1 + \sigma_2} = 0.3357$$

(c) What is the rank of A?

SOLUTION: The rank is the dimension of the column space, which (given the SVD), is the number of non-zero singular values. In this case, the answer is 2.

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

SOLUTION: We must use the reduced form of the SVD. In this case,

$$\Sigma^{-1} = \begin{bmatrix} 1/5.72 & 0 \\ 0 & 1/2.89 \end{bmatrix} \approx \begin{bmatrix} 0.1748 & 0 \\ 0 & 0.3460 \end{bmatrix}$$

and, using some Matlab notation:

$$A^{\dagger} = V(:, 1:2)\Sigma^{-1}U(:, 1:2)^T$$

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

SOLUTION: B^TB is the 2×2 identity. BB^T is the projection matrix to the column space of A (or more generally, to the space spanned by the first two columns of U).

3. 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 $\mathbb{R}^{100}$. 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?

SOLUTION: Since the space has an orthonormal basis,

$$\mathbf{x} = (\mathbf{x} \cdot \mathbf{v}_1)\,\mathbf{v}_1 + (\mathbf{x} \cdot \mathbf{v}_2)\,\mathbf{v}_2 + (\mathbf{x} \cdot \mathbf{v}_3)\,\mathbf{v}_3$$

The coordinate vector for $\mathbf{x}$, $[\mathbf{x} \cdot \mathbf{v}_1, \mathbf{x} \cdot \mathbf{v}_2, \mathbf{x} \cdot \mathbf{v}_3]^T$, is the three-dimensional representation of $\mathbf{x}$. If V is a matrix whose first three columns are $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$, then the low dimensional representation is simply

$$V^T\mathbf{x}$$

4. What is an RBF? Explain it so that a fellow student could follow your reasoning.

   A "Radial Basis Function" is a (nonlinear) function from $\mathbb{R}^n$ to $\mathbb{R}^m$. The action of the RBF on a sample point $\mathbf{x}$ takes the following steps:

   (a) Compute the distance from $\mathbf{x}$ to each of k "centers"- The centers are also vectors in $\mathbb{R}^n$ that are adjustable parameters ("training" an RBF can find the centers). This produces a vector $\mathbf{u} \in \mathbb{R}^k$.

   (b) Apply a nonlinear transfer function ϕ to each element of $\mathbf{u}$ to obtain the vector (still in $\mathbb{R}^k$) $\phi(\mathbf{u})$. The transfer function has to be determined as part of the definition of an RBF.

   (c) Multiply by a matrix of weights, W and add the bias, $\mathbf{b}$. If the desired output is $\mathbb{R}^m$, then W is a matrix that is $m \times k$, and we get:

   $$W\phi(\mathbf{u}) + \mathbf{b} = \mathbf{y}$$

   The weights and bias vector are parameters we find in "training".

5. In Matlab, what does it mean to "train" an RBF? What does it mean to "simulate" an RBF?

   "Training" the RBF means that we will find the set of centers (using OLS) and the weights W and bias $\mathbf{b}$ that "best fits" a given training set of data.

   To "simulate" the network means to produce range elements given certain domain elements. For example, Suppose `Ptrain`, `Ttrain` are patterns and targets for training. Then Matlab will use these to set the centers, weights and bias terms. We would then put the data in `Ptest` to get some outputs- We would compare them to the actual outputs to see how we did.

6. Finish the definition: U is an orthogonal matrix if: U is $n \times n$ and $U^T U = I$. *NOTE: U must be square from its definition...*

7. Finish the definition: Let P be $n \times n$. Then P is a projector if: $P^2 = P$

8. Finish the definition: Let P be $n \times n$. Then P is an orthogonal projector if: P is a projector, and P is symmetric- Altogether, this means $P^2 = P$ and $P^T = P$.

9. If P is a projector, could P be invertible? Explain.

   SOLUTION: If P is invertible, then $P^2 = P$ would imply that $P = I$ (multiply both sides by P^{-1}. Therefore, the only invertible projector is the identity matrix.

10. What is the orthogonal projector to the space spanned by a single vector $\mathbf{a}$?

    SOLUTION:

    $$P = \frac{\mathbf{a}\mathbf{a}^T}{\mathbf{a}^T\mathbf{a}}$$

    (NOTE: The numerator is an $n \times n$ matrix, the denominator is a scalar.

11. Let the matrix X be given below- It has 3 points in $\mathbb{R}^2$.

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

(a) Compute the EDM.

SOLUTION: The full EDM will be 3×3,

$$\begin{bmatrix} 0 & \sqrt{5} & \sqrt{10} \\ \sqrt{5} & 0 & \sqrt{10} \\ \sqrt{5} & \sqrt{10} & 0 \end{bmatrix}$$

(Did you take advantage of the fact that the full EDM has zeros along the diagonal and is symmetric?)

(b) Using all the data as "centers", compute the matrix Φ, where

$$\Phi_{ij} = \phi(\|\mathbf{x}_j - \mathbf{c}_i\|)$$

and $\phi(r) = r^3$.

SOLUTION: Just cube the elements we just found:

$$\begin{bmatrix} 0 & \sqrt{5^3} & \sqrt{10^3} \\ \sqrt{5^3} & 0 & \sqrt{10^3} \\ \sqrt{5^3} & \sqrt{10^3} & 0 \end{bmatrix}$$

(c) What size would Y need to be, and what size would the matrix W be if we solve $W\Phi = Y$ for W? What if we wanted to include a bias (or constant) term?

SOLUTION: The only restriction on Y is that it must contain three vectors. If these vectors are in $\mathbb{R}^m$, then we could take Y to be $m \times 3$. In that case, W would be $m \times 3$ (and Φ is 3×3). If we want to include a bias term, add a last column to W, and add a row of 1's to Φ.

12. What is the Gram-Schmidt algorithm? Give an example using

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix}$$

HINT: When you compute these by hand, you don't need to normalize them.

SOLUTION: We went over the general algorithm in class- We create an orthonormal set whose partial spans are the same in the sense that we maintain the subspace spanned by the first k columns, for $k = 1$ until the end.

In this particular case, we'll compute a set of vectors $\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3$:

- $\mathbf{w}_1 = \mathbf{v}_1$
- $\mathbf{w}_2 = \mathbf{v}_2 - \frac{\mathbf{v}_2 \cdot \mathbf{w}_1}{\mathbf{w}_1 \cdot \mathbf{w}_1} \mathbf{w}_1 = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$

- $\mathbf{w}_3 = \mathbf{v}_3 - \frac{\mathbf{v}_3 \cdot \mathbf{w}_1}{\mathbf{w}_1 \cdot \mathbf{w}_1}\mathbf{w}_1 - \frac{\mathbf{v}_3 \cdot \mathbf{w}_2}{\mathbf{w}_2 \cdot \mathbf{w}_2}\mathbf{w}_2 = \begin{bmatrix} 1/6 \\ -1/3 \\ -1/6 \end{bmatrix}$

13. Given the vector $\mathbf{x} = [1, 1, 1]^T$, which vector

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix}$$

"best lies in the direction of $\mathbf{x}$" (in the sense of the subspace spanned by the vector, as in the OLS).

SOLUTION: We compute $\cos^2(\theta_j) = \frac{(\mathbf{x}^T \mathbf{v}_j)^2}{\|\mathbf{x}\|^2 \|\mathbf{v}_j\|^2}$, where θ_j is the angle between $\mathbf{x}$ and $\mathbf{v}_j$. In order, these values are:

$$\frac{(1+1)^2}{2 \cdot 3} = \frac{4}{6} = \frac{2}{3} \qquad \frac{(3+1+1)^2}{11 \cdot 3} = \frac{25}{33} \approx 0.758 \qquad \frac{(2-1+3)^2}{14 \cdot 3} = \frac{16}{42} \approx 0.381$$

Therefore, column 2 is closest.

14. Let the matrix X store 4 points in $\mathbb{R}^2$, below:

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

(a) Write the system of equations to find the polynomial, $p(x)$, that interpolates the data.

SOLUTION: Let the cubic function be given by $c_4 x^3 + c_3 x^2 + c_2 x + c_1 = y$. Then the system of equations (in matrix form) would be:

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 1 \\ 8 & 4 & 2 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} c_4 \\ c_3 \\ c_2 \\ c_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 2 \\ 1 \end{bmatrix}$$

(b) If we're looking for a function that maps the elements of the first row to the second row, what would the form of the RBF be (using the first three points as centers)? Write it out-

SOLUTION: In this case, x, y are scalars, and:

$$y = w_1 \phi(|x-1|) + w_2 \phi(|x+1|) + w_3 \phi(|x-2|) + b$$

Answers to the extra practice questions.

1. What is the reduced SVD for the matrix A? The matrix is one dimensional, so it is:

$$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} 2\,[1\ 0] \quad \Rightarrow \quad A^{\dagger} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \frac{1}{2}\,[1\ 0\ 0] = \begin{bmatrix} 1/2 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$