Math Homework 2

Math 140: Industrial Math

Scott Oslund
Aleks Rezler
Renier Briseno
Zihao Li

January 23, 2023

1)

Solution:

We want to show that $x_\perp$ is orthogonal to v. We know that two vectors are orthogonal if their dot product is zero. Therefore we have to show that

$$x_\perp \cdot v = 0$$

To do this, we can first apply the given definition of $x_\perp$

$$x_\perp \cdot v$$

$$(x - x_\parallel) \cdot v$$

We know the dot product is distributive, thus we can write

$$x \cdot v - x_\parallel \cdot v$$

We can now apply the given definition of $x_\parallel$

$$x \cdot v - (vv^T x) \cdot v$$

Next, in our equation, we can note that $v^T x$ will be a scalar since we have a row vector being multiplied by a column vector. This operation is equivalent to taking the dot product. Then, we multiply this scalar by the vector v. scalar vector multiplication is commutative we can write rewrite the above as

$$x \cdot v - (v^T x v) \cdot v$$

$$x \cdot v - v^T x v \cdot v$$

$v \cdot v$ is simply the norm of v. In this case v is a unit vector, so this equals 1. Therefore we can simplify to

$$x \cdot v - v^T x$$

And lastly, $v^T x$ is clearly equivalent to the dot product of v and x since it multiplies every element of v with the corresponding element of x and sums them together.

$$x \cdot v - x \cdot v = 0$$

Therefore we have shown

$$x_\perp \cdot v = 0$$

and therefore $x_\perp$ and v are orthogonal.

The orthogonality of these vectors is useful because in PCA we want each Principal component to be orthogonal to the other ones. If we did not make them orthogonal then there would be correlation between the principal components and redundancy. This property says that once we create the first principal component and project the data points onto the principal component, the remaining variation in the datapoint not captured by the first principal component will be orthogonal to the principal component. This means the following principal components can also be orthogonal to the first principal component.

2)

Solution: To solve this problem, we want to show

$$||v||^2 = \sum_{i=1}^{N} v_i^2$$

First, we know that $||v||^2 = \langle v, v \rangle$.

$$\langle v, v \rangle$$

Next, we can apply the given definition of v

$$\langle \sum_{i=1}^{N} v_i b_i, \sum_{i=1}^{N} v_i b_i \rangle$$

Next, we know the inner product is distributive so we can write:

$$\langle v_1 b_1 + v_2 b_2 ... + v_N b_N, v_1 b_1 + v_2 b_2 ... + v_N b_N \rangle$$

$$\begin{array}{c}
\langle v_1 b_1, v_1 b_1 \rangle + \langle v_1 b_1, v_2 b_2 \rangle ... + \langle v_1 b_1, v_N b_N \rangle + \\
\langle v_2 b_2, v_1 b_1 \rangle + \langle v_2 b_2, v_2 b_2 \rangle ... + \langle v_2 b_2, v_N b_N \rangle + \\
\vdots \qquad\qquad\qquad\qquad\qquad\qquad \vdots \\
\langle v_N b_N, v_1 b_1 \rangle + \langle v_N b_N, v_2 b_2 \rangle ... + \langle v_N b_N, v_N b_N \rangle
\end{array}$$

We know that v_i is a scalar since it is defined as $\langle v, b_i \rangle$. We also know that b_i is an orthonormal basis. Therefore, $\langle v_i b_i, v_j b_j \rangle$ will always be zero in the case when $i \neq j$ since these vectors will be orthogonal. Therefore this all simplifies to

$$\langle v_1 b_1, v_1 b_1 \rangle + \langle v_2 b_2, v_2 b_2 \rangle ... + \langle v_N b_N, v_N b_N \rangle$$

or, equivalently

$$\sum_{i=1}^{N} \langle v_i b_i, v_i b_i \rangle$$

Since as we noted before v_i is a scalar, we can move both of these out of the inner product.

$$\sum_{i=1}^{N} v_i^2 \langle b_i, b_i \rangle$$

And lastly, $\langle b_i, b_i \rangle = 1$ according to the Kronecker delta since these are the same vectors. This leaves us with

$$\sum_{i=1}^{N} v_i^2$$

Therefore we have shown that

$$||v||^2 = \sum_{i=1}^{N} v_i^2$$

7)

Solution:

In order to perform PCA on this data, we must first note it's matrix form.

$$w = \left\{ \begin{pmatrix} -1 \\ -2 \end{pmatrix}, \begin{pmatrix} 1 \\ 3 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 0.5 \end{pmatrix} \right\} = \begin{bmatrix} -1 & -2 \\ 1 & 3 \\ 0 & 1 \\ 2 & 0.5 \end{bmatrix}$$

Now that we have our matrix w, we now must make sure our data is centered. To do this, we can simply subtract off the mean from each dimension. Note,

$$w1_{\text{avg}} = \frac{-1 + 1 + 0 + 2}{4} = 0.5$$

$$w2_{\text{avg}} = \frac{-2 + 3 + 1 + 0.5}{4} = 0.625$$

Subtracting w1 element-wise from the first column and w2 from the second yields,

$$w = \begin{bmatrix} -1.5 & -2.625 \\ 0.5 & 2.375 \\ -0.5 & 0.375 \\ 1.5 & -0.125 \end{bmatrix}$$

Our next step is to find the covariance matrix. Formally, it is given by this formula,

$$cov(X, Y) = \frac{1}{n} \sum_{i=1}^{n} (x - \bar{x})(y - \bar{y})$$

where $\bar{x}$ and $\bar{y}$ is the sample mean. Note that multiplying w by its transpose would usually give us a Sum of Squares and Cross Products Matrix (a symmetric matrix in which we can take advantage of its properties!). Furthermore, since we have already subtracted the mean off, we would end up with a multiple of a Covariance Matrix. Dividing by the number of rows would give,

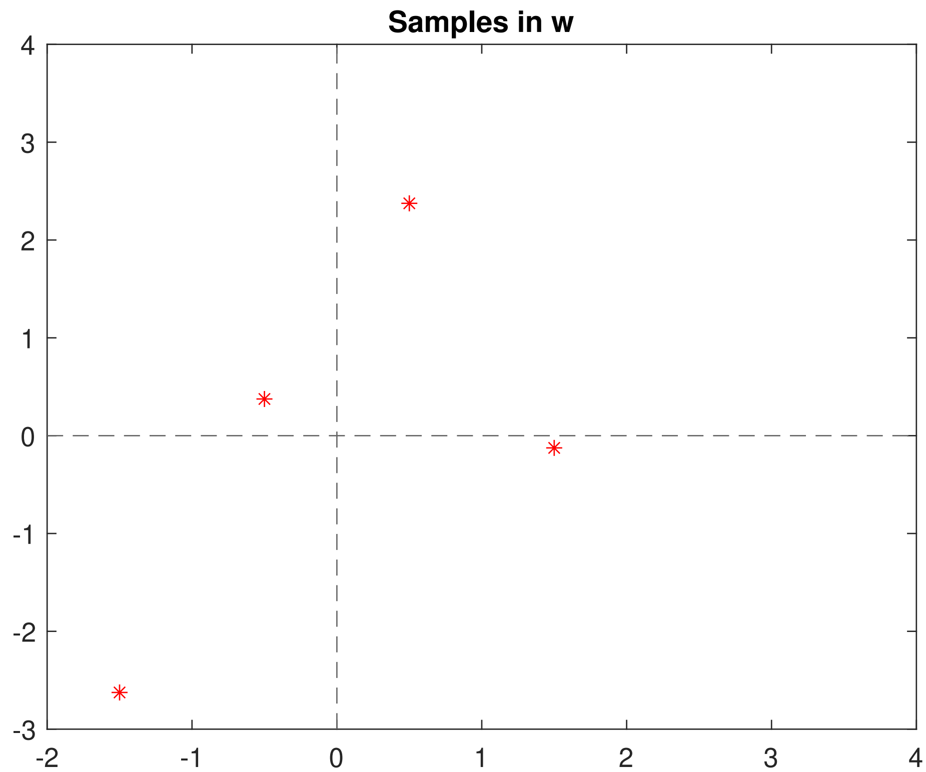
$$\Sigma = \begin{bmatrix} 2.2852 & -1.7461 & -0.0586 & -0.4805 \\ -1.7461 & 1.4727 & 0.1602 & 0.1133 \\ -0.0586 & 0.1602 & 0.0977 & -0.1992 \\ -0.4805 & 0.1133 & -0.1992 & 0.5664 \end{bmatrix}$$

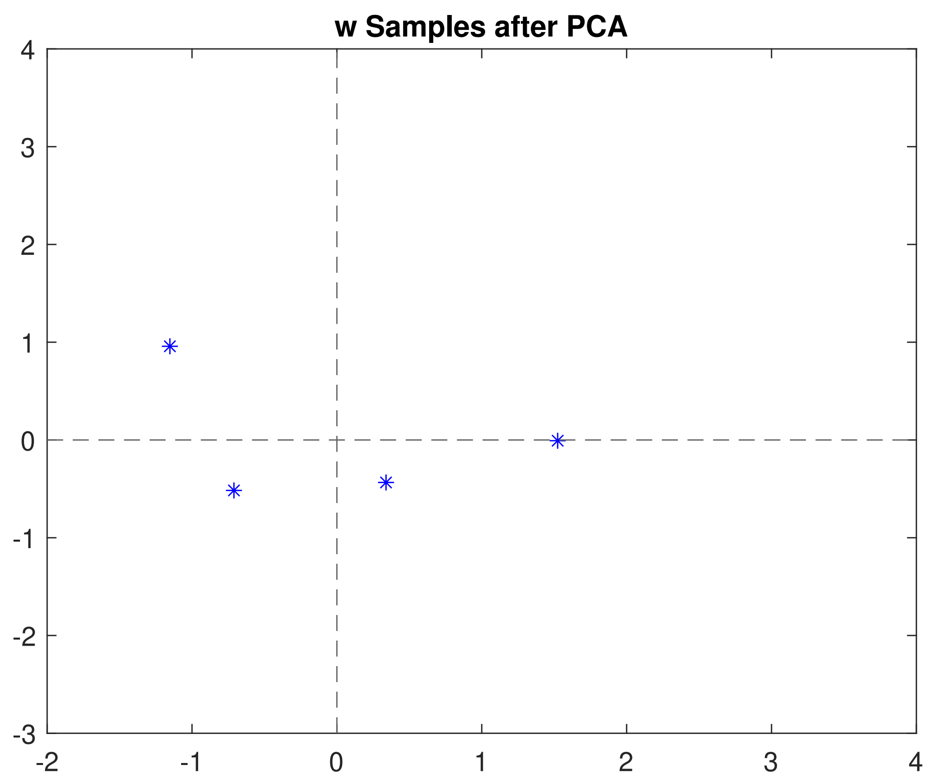
Using a Matlab script, we found our eigenvectors and their eigenvalues to be:

$$\lambda = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0.6833 & 0 \\ 0 & 0 & 0 & 3.7385 \end{bmatrix} \quad E = \begin{bmatrix} 0.4250 & 0.4396 & = 0.1350 & -0.7797 \\ 0.5795 & 0.4154 & -0.3457 & 0.6100 \\ -0.6954 & 0.6148 & -0.3706 & 0.0318 \\ 0 & 0.5061 & 0.8514 & 0.1379 \end{bmatrix}$$

Where each column is the eigenvector of the respective eigenvalue of the matrix on the left.

Clearly, we can reduce dimensionality. Furthermore, we have the choice of keeping both 0 eigenvectors, or discarding the least significant. The contribution of the largest eigenvalue is roughly 84%. In the interest of keeping the total information above 90%, we can keep both of these principal components. Finally, to transform the samples into our new subspace, we can concatenate our ordered eigenvectors into a matrix, take its transpose(4x4), and multiply by the original matrix(4x2) to get our new transformed data set. We can see the results from running the Matlab code.


Samples in w
4
3
2
1
0
-1
-2
-3
-2
-1
0
1
2
3
4


w Samples after PCA
4
3
2
1
0
-1
-2
-3
-2
-1
0
1
2
3
4

Solution: Below is the Matlab code used to perform the PCA on number 7 and make the graphs.

```
clear;
w = [-1,-2;
     1, 3;
     0, 1
     2, 0.5]
% might need to standardize
w(:,1) = w(:,1)-sum(w(:,1))/4;
w(:,2) = w(:,2)-sum(w(:,2))/4
% Plotting the samples in w
figure(1)
plot(w(:,1),w(:,2), "r*")
axis([-2, 4, -3, 4])
xline(0, "LineStyle", "--")
yline(0, "LineStyle", "--")
title("Samples in w")
% getting transpose
w_t = w.';
% covariance matrix
sigma = (w*w_t)./4
% eigs of sigma
[V, D] = eig(sigma)
% sorting eigenvectors
E1 = V(:,4)
E2 = V(:,3);
E3 = V(:,1);
E4 = V(:,2);
% checking that E(1:4) are in fact eigenvectors of sigma
c = sigma*E1
d = D(4,4)*E1
E = [V(:,4), V(:,3), V(:,1), V(:,2)];
E_T = E.'
Final = E_T*w
% Plotting the principal components
figure(2)
plot(Final(:,1),Final(:,2), "b*")
axis([-2, 4, -3, 4])
xline(0, "LineStyle", "--")
yline(0, "LineStyle", "--")
title("w Samples after PCA")
hold off
```