Exam 1 Review Solutions (Spr 25)

Be sure to read over and understand the solutions to quiz as well as the review sheets.

1. Given data $\{x_1, \ldots, x_n\}$

   - Sample mean: $\bar{x} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} x_i$
   - Sample variance: $s^2 = \dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}(x_i - \bar{x})^2$
   - Given a second data set $\{y_1, y_2, \ldots, y_p\}$ (note that order matters), then

     Covariance: $\dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$

     And this is a scaled dot product between these two "vectors" of mean subtracted data:

     $\dfrac{1}{n-1}(\mathbf{x} - \bar{x}) \cdot (\mathbf{y} - \bar{y})$

     Correlation: Scaled covariance (scaled so that the two vectors have unit size):

     $$\frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

     (Notice that the $1/(n-1)$ term cancels top and bottom). Given two vectors of mean-subtracted data, this reduces to:

     $$\frac{(\mathbf{x} - \bar{x}) \cdot (\mathbf{y} - \bar{y})}{\|\mathbf{x} - \bar{x}\| \, \|\mathbf{y} - \bar{y}\|} = \cos(\theta)$$

     where θ is the angle between the vectors $\mathbf{x} - \bar{x}$ and $\mathbf{y} - \bar{y}$.

2. For the covariance matrix, if we have n dimensional data, then the covariance matrix is $n \times n$, where the $(i, j)^{\text{th}}$ entry is the covariance between the p values in dimension i against the p values in dimension j. If X is $n \times p$, then these vectors are stored as the rows of X. Be sure to mean subtract the matrix first (in this case, the mean is a column in $\mathbb{R}^n$). If we define $\hat{X}$ as the mean-subtracted matrix, then the covariance matrix is given as

   $$C = \frac{1}{p-1}\hat{X}\hat{X}^T$$

   We would note that it is possible that the matrix X is $p \times n$, where the data in dimension i is the i^{th} column. In that case, the mean would be a row vector. Define $\hat{X}$ as the mean-subtracted matrix, and then the covariance is

   $$C = \frac{1}{p-1}\hat{X}^T\hat{X}$$

   In both cases, C is $n \times n$, so you can see that it's important to pay attention to the dimensions of objects you're working with!

3. 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.

SOLUTION: Let the vectors for G be denoted by $\boldsymbol{u}_1, \boldsymbol{u}_2$ (and notice that these are orthogonal to each other). Let $\hat{\boldsymbol{x}}$ be the projection to G:

$$\hat{\boldsymbol{x}} = \frac{\boldsymbol{u}_1 \cdot \boldsymbol{x}}{\boldsymbol{u}_1 \cdot \boldsymbol{u}_1}\boldsymbol{u}_1 + \frac{\boldsymbol{u}_2 \cdot \boldsymbol{x}}{\boldsymbol{u}_2 \cdot \boldsymbol{u}_2}\boldsymbol{u}_2 = -\frac{3}{14}\boldsymbol{u}_1 + \frac{3}{10}\boldsymbol{u}_2 \approx \begin{bmatrix} 0.6857 \\ -0.9429 \\ 0.4286 \end{bmatrix}$$

The distance to the plane is $\|\boldsymbol{x} - \hat{\boldsymbol{x}}\|$, which in this case is approximately 1.8593.

(NOTE: On an exam, the numbers would come out in a way that you could do them by hand. In this example, note what computations we made, but the arithmetic on the exam would work out nicer.)

4. 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)?

SOLUTION:

$$\boldsymbol{x} = 3\boldsymbol{b}_1 - \boldsymbol{b}_2 = \begin{bmatrix} 0 \\ -1 \\ 2 \end{bmatrix}$$

5. 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 $[\boldsymbol{x}]_{\mathcal{B}}$?

SOLUTION: In this case, we have to compute the coordinates- That is, we have to solve the system:

$$\begin{bmatrix} 6 & 1 \\ 1 & -2 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} 3 \\ -1 \end{bmatrix}$$

Solving it is easiest if you happen to recall the formula for inverting a 2×2 matrix, but if you don't, you can also construct the appropriate augmented matrix and row reduce:

$$\boldsymbol{c} = -\frac{1}{13} \begin{bmatrix} -2 & -1 \\ -1 & 6 \end{bmatrix} \begin{bmatrix} 3 \\ -1 \end{bmatrix} = \begin{bmatrix} 5/13 \\ 9/13 \end{bmatrix}$$

(NOTE: I'll try to make the arithmetic work out nicely on the exam- the important point is to recall what operations to perform).

6. 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 $\boldsymbol{a}$.

   SOLUTION:
$$A = \frac{\boldsymbol{a}\boldsymbol{a}^T}{\boldsymbol{a}^T\boldsymbol{a}} = \frac{1}{10}\begin{bmatrix} 1 & 3 \\ 3 & 9 \end{bmatrix}$$

   Check:
$$A\mathbf{x} = \frac{\boldsymbol{a}\boldsymbol{a}^T}{\boldsymbol{a}^T\boldsymbol{a}}\boldsymbol{x} = \frac{1}{\boldsymbol{a}^T\boldsymbol{a}}\boldsymbol{a}(\boldsymbol{a}^T\boldsymbol{x}) = \frac{\boldsymbol{a}^T\boldsymbol{x}}{\boldsymbol{a}^T\boldsymbol{a}}\boldsymbol{a}$$

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

   SOLUTION: First, we should note that the plane needs to go through the origin before we can consider it a two dimensional subspace (otherwise, we would need more information). With that, there are a couple of ways to compute this- One way is that we need a spanning set for the plane. The plane is the set of (x, y, z) such that $x+y+z = 0$, which is a linear system that we can solve. In fact, the "matrix": $[1\ 1\ 1|0]$ is alread in RREF.

   Solving the system, we have two free variables, y, z, and we solve the system as usual. Put the two vectors as columns in a matrix A:
$$\begin{array}{lll} x & = -y & -z \\ y & = y & \\ z & = & z \end{array} \quad \Rightarrow \quad A = \begin{bmatrix} -1 & -1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

   Now the projection matrix is:
$$P = A(A^T A)^{-1} A^T$$

   which can be computed as (you can leave it unsimplified):
$$\begin{bmatrix} -1 & -1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 2/3 & -2/3 \\ -1/3 & 2/3 \end{bmatrix} \begin{bmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix}$$

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

   (a) $A = \begin{bmatrix} 5 & -1 \\ 3 & 1 \end{bmatrix}$

   The characteristic equation is $\lambda^2 - 6\lambda + 8 = 0$, so $\lambda = 2, 4$.

   For $\lambda = 2$, we solve $(5 - 2)v_1 - v_2 = 0$, and so we use
$$\mathbf{v} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$$

Similarly, for $\lambda = 4$, we solve $(5-4)v_1 - v_2 = 0$, so we use

$$\mathbf{v} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

(Note: The "shortcut" for finding $\boldsymbol{v}$ only works here because our matrix A is 2×2. Otherwise, we would have to row reduce the matrix $A - \lambda I$).

(b) $A = \begin{bmatrix} -2 & 1 \\ 1 & -2 \end{bmatrix}$

The characteristic equation is $\lambda^2 + 4\lambda + 3 = 0$, so $\lambda = -1, -3$.
For $\lambda = -1$, we solve $(-2+1)v_1 + v_2 = 0$, and so we use

$$\mathbf{v} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Similarly, for $\lambda = -3$, we solve $(-2+3)v_1 + v_2 = 0$, so we use

$$\mathbf{v} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

(Note: The "shortcut" for finding $\boldsymbol{v}$ only works here because our matrix A is 2×2. Otherwise, we would have to row reduce the matrix $A - \lambda I$).

9. (Referring to the previous exercise) We could've predicted that the eigenvalues of the second matrix would be real, and that the eigenvectors would be orthogonal. Why?

   SOLUTION: The second matrix is symmetric, so by the Spectral Decomposition Theorem, it has two real eigenvalues and orthogonal eigenvectors.

10. Compute the SVD of the matrix A, and the pseudoinverse of A, given the matrix below:

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

SOLUTION: A couple of matrix computations first:

$$A^T A = \begin{bmatrix} 5 & 0 \\ 0 & 0 \end{bmatrix} \qquad AA^T = \begin{bmatrix} 1 & 2 & 0 \\ 2 & 4 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

From this, we see that $\lambda_1 = 5$, so $\sigma_1 = \sqrt{5}$, and the remaining singular values (and eigenvalues) are 0. The eigenvector for $\lambda = 5$ in $A^T A$ is $(1, 0)$ The (unscaled) corresponding eigenvector of AA^T is given by

$$A\boldsymbol{v}_1 = \begin{bmatrix} 1 & 0 \\ 2 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} \quad \Rightarrow \quad \boldsymbol{u}_1 = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}$$

Now, the other eigenvector $\boldsymbol{v}_2$ is orthogonal to $\boldsymbol{v}_1$, so we'll take it as $(0,1)$. The eigenspace of AA^T for $\lambda = 0$ will be two dimensional, so to find a basis for that, we row reduce $AA^T - \lambda I$, which is just AA^T. That's easy:

$$AA^T = \begin{bmatrix} 1 & 2 & 0 \\ 2 & 4 & 0 \\ 0 & 0 & 0 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \rightarrow \begin{array}{ll} x_1 & = -2x_2 \\ x_2 & = x_2 \\ x_3 & = \qquad x_3 \end{array} \rightarrow \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Therefore, the full SVD is given by:

$$\begin{bmatrix} 1 & 0 \\ 2 & 0 \\ 0 & 0 \end{bmatrix} = U\Sigma V^T = \begin{bmatrix} 1/\sqrt{5} & -2\sqrt{5} & 0 \\ 2/\sqrt{5} & 1\sqrt{5} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{5} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}^T$$

Remember that the pseudoinverse uses the **reduced** SVD, so it will be:

$$A^\dagger = V(:,1)\Sigma^{-1}(1,1)U(:,1)^T = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \frac{1}{\sqrt{5}} \begin{bmatrix} 1/\sqrt{5} & 2/\sqrt{5} & 0 \end{bmatrix} = \begin{bmatrix} 1/5 & 2/5 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

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

SOLUTION:

$$P = \frac{\boldsymbol{x}\boldsymbol{x}'}{\boldsymbol{x}'\boldsymbol{x}} = \frac{1}{3} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

12. 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.

SOLUTI)ON:

- The first part of the question is incorrect. It asks for the coordinates of $\boldsymbol{x}$ with respect to the columns of U, but $\boldsymbol{x}$ is not contained in the columnspace of U (notice that the two columns both have zero in the third spot, but $\boldsymbol{x}$ does not).

  What was meant was that we want to find the coordinates of the projection of $\boldsymbol{x}$ into the columnspace of U. In that case,

$$[\text{Proj}(\boldsymbol{x})]_U = U^T\boldsymbol{x} = \begin{bmatrix} 4/\sqrt{2} \\ 2/\sqrt{2} \end{bmatrix}$$

- The projection of $\boldsymbol{x}$ into the columnspace of U is given by $UU^T\boldsymbol{x}$, or $(1,3,0)$.

- The distance between $\boldsymbol{x}$ and its projection is easy to compute this time: $\|\boldsymbol{x} - (1,3,0)\| = \|(0,0,2)\| = 2$

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

    Let $\boldsymbol{x}$ be any vector in the null space. We will show that the dot product between $\boldsymbol{x}$ and any row of A is zero.

    Consider $A\boldsymbol{x}$ in terms of the rows of A, so let $\boldsymbol{r}_i$ be the i^{th} row of A. Then, assuming A is $m \times n$,

$$A\boldsymbol{x} = \begin{bmatrix} \boldsymbol{r}_1 \\ \boldsymbol{r}_2 \\ \vdots \\ \boldsymbol{r}_m \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} \boldsymbol{r}_1 \cdot \boldsymbol{x} \\ \boldsymbol{r}_2 \cdot \boldsymbol{x} \\ \vdots \\ \boldsymbol{r}_m \cdot \boldsymbol{x} \end{bmatrix} = \vec{0}$$

    Therefore, the dot product of every row of A with $\boldsymbol{x}$ is 0, so every row is orthogonal to $\boldsymbol{x}$. Since $\boldsymbol{x}$ was arbitrary in the null space, then every row is orthogonal to every element in the null space.

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

    SOLUTION: One way to do this is using the determinants. If the matrices are square, then $\det(AB) = \det(A)\det(B)$. Now,

$$\det(X^{-1}AX - \lambda I) = \det(X^{-1}AX - \lambda X^{-1}X) = \det(X^{-1}(A - \lambda I)X) =$$

$$\det(X^{-1})\det(A - \lambda I)\det(X) = \det(A - \lambda I)$$

    Therefore, the characteristic equation for A and the characteristic equation for $X^{-1}AX$ are the same (and so they have the same eigenvalues).

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

    SOLUTION: There are a couple of different ways. One way: If $\boldsymbol{a}$ is a column vector mean, $\boldsymbol{b}$ is a row vector mean, and c is the grand mean (the mean over all elements of the matrix), then we double center by taking:

$$X - a - b + c$$

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

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

    True: The rank is the dimension of the column space, but that is always equal to the dimension of the row space.

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

    False: A correlation coefficient of 1 means that the data is linearly related (there is a linear relationship between the data).

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

False: A correlation of 0 means that there is no linear relationship between the data ("uncorrelated"), but there could be a nonlinear relationship between the data.

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

With the added assumption that U has orthonormal columns, then "True". Otherwise, false. (As written, it would be false).

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

False. If U has orthonormal columns, then UU^T is the projection onto the column space of U, which may not be I. Of course, if U doesn't have o.n. columns, it would be false.

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

With the added assumption that A is square, then this is true, since "A not invertible" would mean that $\det(A - 0I) = 0$.

(g) Let

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

Then the rank of AA^T is 2.

True. The rank of A is the same as the rank of AA^T (and $A^T A$).

17. Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ be the normalized eigenvectors of $A^T A$, where A is $m \times n$.

(a) Show that if λ_i is a non-zero eigenvalue of $A^T A$, then it is also a non-zero eigenvalue of AA^T.

SOLUTION: If λ_i, $\boldsymbol{v}_i$ is the eigenvalue/eigenvector pair for $A^T A$, then by definition,

$$A^T A \boldsymbol{v}_i = \lambda_i \boldsymbol{v}_i$$

Multiply both sides by matrix A:

$$AA^T A \boldsymbol{v}_i = \lambda_i A \boldsymbol{v}_i \quad \Rightarrow \quad AA^T \boldsymbol{u}_i = \lambda_i \boldsymbol{u}_i$$

where $\boldsymbol{u}_i = A\boldsymbol{v}_i$. Therefore, λ_i is an eigenvalue of AA^T.

(b) True or false? The eigenvectors form an orthogonal basis of $\mathbb{R}^n$.

True. This is a result of the Spectral Theorem.

(c) Show that, if $\mathbf{x} \in \mathbb{R}^n$, then the i^{th} coordinate of $\mathbf{x}$ (with respect to the eigenvector basis) is $\mathbf{x}^T \mathbf{v}_i$.

SOLUTION: To show this directly, we can start with

$$\boldsymbol{x} = c_1 \boldsymbol{v}_1 + c_2 \boldsymbol{v}_2 + \ldots + c_n \boldsymbol{v}_n$$

Now dot both sides with $\boldsymbol{v}_i$. By the last question, the eigenvectors are orthonormal, so

$$\boldsymbol{x} \cdot \boldsymbol{v}_i = 0 + 0 + c_i + 0 + \ldots + 0$$

which is what we wanted to show.

(d) Let $\alpha_1, \ldots, \alpha_n$ be the coordinates of $\mathbf{x}$ with respect to $\mathbf{v}_1, \ldots, \mathbf{v}_n$.
Show that

$$\|\mathbf{x}\|_2 = \alpha_1^2 + \alpha_2^2 + \ldots + \alpha_n^2$$

I'll allow you to show it just using just two vectors, $\mathbf{v}_1, \mathbf{v}_2$.
TYPO: The RHS should be $\|\boldsymbol{x}\|^2$. In that case,

$$\|\boldsymbol{x}\|^2 = \boldsymbol{x} \cdot \boldsymbol{x} = (\alpha_1 \boldsymbol{v}_1 + \alpha_2 \boldsymbol{v}_2) \cdot (\alpha_1 \boldsymbol{v}_1 + \alpha_2 \boldsymbol{v}_2)$$

Expanding this expression, $\boldsymbol{v}_i \cdot \boldsymbol{v}_j = 0$ if $i \neq j$, so we're just left with:

$$\alpha_1^2 \boldsymbol{v}_1 \cdot \boldsymbol{v}_1 + \alpha_2^2 \boldsymbol{v}_2 \cdot \boldsymbol{v}_2$$

so if we have orthonormal vectors, this reduces to $\alpha_1^2 + \alpha_2^2$.
(The general case is very similar).

(e) Show that $A\mathbf{v}_i \perp A\mathbf{v}_j$
SOLUTION: Show that the dot product is 0:

$$(A\boldsymbol{v}_i) \cdot (A\boldsymbol{v}_j) = (A\boldsymbol{v}_i)^T (A\boldsymbol{v}_j) = \boldsymbol{v}_i^T A^T A \boldsymbol{v}_j = \boldsymbol{v}_i \lambda_j \boldsymbol{v}_j = \lambda_j \boldsymbol{v}_i^T \boldsymbol{v}_j = 0$$

(f) Show that $A\mathbf{v}_i$ is an eigenvector of AA^T.
SOLUTION: See part (a). Same argument.

18. 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.

SOLUTION: This is repeating what we had done in class. I'm using t_i for the i^{th} "target" rather than y_i, but otherwise, this is straightforward.

We want to find m, b so that $t_i = mx_i + b$ for each $i = 1, 2, \ldots, p$. This gives an error in m, b:

$$E(m, b) = \sum_{i=1}^{p} (t_i - (mx_i + b))^2$$

Using calculus, we can take the partial derivatives of E with respect to m, b, and we wind up with the solution to the normal equations.

$$\frac{\partial E}{\partial m} = \sum_{i=1}^{p} 2(t_i - (mx_i + b)))(-x_i) = 0$$

Distribute the sum through to get:

$$m\sum_{i=1}^{p} x_i^2 + b\sum_{i=1}^{p} x_i = \sum_{i=1}^{p} x_i t_i$$

Similarly,

$$\frac{\partial E}{\partial b} = \sum_{i=1}^{p} 2(t_i - (mx_i + b))(-1)$$

Distribute the sum through, and note that $\sum_{i=1}^{p} 1 = p$ so that

$$m\sum_{i=1}^{p} x_i + bp = \sum_{i=1}^{p} t_i$$

For the linear algebra version, let the design matrix A and the unknown vector $\boldsymbol{c}$ be given by

$$A = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_p & 1 \end{bmatrix} \boldsymbol{c} = \begin{bmatrix} m \\ b \end{bmatrix}$$

Then we want to solve $A\boldsymbol{c} = \boldsymbol{t}$. Multiplying both sides by A^T, we get the normal equation:

$$A^T A\boldsymbol{c} = A^T\boldsymbol{t}$$

Multiplying these out, we get the same system of equations as we did when setting the partial derivatives equal to zero.

19. Given data:

x	-1	0	1
y	2	1	1

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

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

(b) Compute the normal equations.

$$\begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} -1 \\ 4 \end{bmatrix}$$

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

$$m = -1/2, \qquad b = 4/3$$

20. Use the data in Exercise (19) 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?)

SOLUTION: The matrix equation is now:

$$\begin{bmatrix} 1 & -1 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}$$

This matrix is invertible, so we get an actual unique solution to this system. Inverting the matrix, we get that

$$a = 1/2 \qquad b = -1/2 \qquad c = 1$$

21. 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 $\mathbf{x}\mathbf{x}^T$). HINT: SVD

SOLUTION:

- The matrix $\boldsymbol{x}\boldsymbol{x}^T$ is $\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$
- One eigenvector is $\boldsymbol{x}$, since

$$(\boldsymbol{x}\boldsymbol{x}^T)\boldsymbol{x} = (\boldsymbol{x}^T\boldsymbol{x})\boldsymbol{x}$$

so the corresponding eigenvalue is $\boldsymbol{x}^T\boldsymbol{x} = 6$.

The other eigenvalues are zero, so we would just need to row reduce $\mathbf{x}\mathbf{x}^T$, which is easy to do:

$$\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \rightarrow \begin{array}{lll} x_1 & = -2x_2 & -x_3 \\ x_2 & = x_2 & \\ x_3 & = & x_3 \end{array} \rightarrow \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$

These are the eigenvectors for $\lambda = 0$. You might notice that these are not orthogonal. If we were doing an SVD, then we would need to do Gram-Schmidt on these two vectors to get an orthonormal set. However, you might also notice that both vectors are orthogonal to the other eigenvector, $(1, 2, 1)$.

22. 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 $\mathbf{x}$.
   SOLUTION:

$$\bar{\boldsymbol{y}} = m\bar{\boldsymbol{x}} + b$$

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

SOLUTION:

$$\mathrm{Cov}(\boldsymbol{x}+a, \boldsymbol{y}+b) = \frac{1}{n-1}\sum_{i=1}^{n}((\boldsymbol{x}+a)-(\bar{\boldsymbol{x}}+a))((\boldsymbol{y}+b)-(\bar{\boldsymbol{y}}+b)) =$$

$$\frac{1}{n-1}\sum_{i=1}^{n}(\boldsymbol{x}-\bar{\boldsymbol{x}})(\boldsymbol{y}-\bar{\boldsymbol{y}})) = \mathrm{Cov}(\boldsymbol{x}, \boldsymbol{y})$$

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

SOLUTION:

$$s_{xy} = \mathrm{Cov}(\boldsymbol{x}, m\boldsymbol{x}+b) = \frac{1}{n-1}\sum_{i=1}^{n}((\boldsymbol{x}-\bar{\boldsymbol{x}})(m\boldsymbol{x}+b)-(m\bar{\boldsymbol{x}}+b)) =$$

$$m\frac{1}{n-1}\sum_{i=1}^{n}(\boldsymbol{x}-\bar{\boldsymbol{x}})^2 = ms_x^2$$

For the correlation, first consider the variance of $\boldsymbol{y}$:

$$s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}((m\boldsymbol{x}+b)-(m\bar{\boldsymbol{x}}+b))^2 = m^2 s_x^2 \quad \Rightarrow \quad s_y = |m|s_x$$

$$r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{ms_x^2}{s_x|m|s_x} = \frac{m}{|m|}$$

This expression is equal to 1 if $m > 0$ and -1 if $m < 0$ (it's called *signum(m)*).

23. 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?

    SOLUTION: The low dimensional representation is in $\mathbb{R}^3$, and is given by the coordinates of $\boldsymbol{x}$:

$$[\boldsymbol{x}]_W = (\boldsymbol{x}\cdot\boldsymbol{v}_1, \boldsymbol{x}\cdot\boldsymbol{v}_2, \boldsymbol{x}\cdot\boldsymbol{v}_3)$$

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.

    SOLUTION: $A = [1\ 3\ 4]$.

    Notation: Parenthesis will denote a column vector, like $(1, 2)$ is the same as $[1, 2]^T$. This notation was also used in Lay's linear algebra text.

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

    SOLUTION: The rank of A is 1, so the dimension of the row space and column space are both 1. The null space is therefore 2 dimensional, and the null space of A^T is zero dimensional.

(b) Find bases for the four fundamental subspaces.

SOLUTION:

- Row space: $\{(1,3,4)\}$.
- Null space:

$$\begin{array}{lll} x & = -3y & -4z \\ y & = y & \\ z & = & z \end{array} \qquad \Rightarrow \qquad \left\{ \begin{bmatrix} -3 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -4 \\ 0 \\ 1 \end{bmatrix} \right\}$$

- Column space: $\{1\}$
- Null space of A^T (just the zero vector)

(c) Find a solution to the equation with at least 2 zeros.

$$x = 1, y = 0, z = 0$$

(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.

SOLUTION: This is actually the projection onto a vector, since the column space of A is the span of one vector (call it $\boldsymbol{a}$)). Therefore,

$$P = \frac{1}{\boldsymbol{a}^T\boldsymbol{a}}\boldsymbol{a}\boldsymbol{a}^T = \frac{1}{1^2+3^2+4^2} \begin{bmatrix} 1 & 3 & 4 \\ 3 & 9 & 12 \\ 4 & 12 & 16 \end{bmatrix}$$

25. (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
```

SOLUTION: Before answering, the size of A is the same as the size of S in the full SVD, so A is 5×3, and the mapping $A\boldsymbol{x}$ goes from $\mathbb{R}^3$ to $\mathbb{R}^5$. Further, we see (from S) that the rank of A is 2.

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

SOLUTION: The null space is in $\mathbb{R}^3$, and is one dimensional- It is the last column of V.

For the row space of A?

SOLUTION: The row space is in $\mathbb{R}^3$, and a basis would use the first two columns of V.

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

SOLUTION: In class, we normalized the eigenvalues rather than the singular values, but we could do the same thing to the singular values. That would be:

$$\frac{5.72}{5.72 + 2.89}, \qquad \frac{2.89}{5.72 + 2.89}$$

(c) What is the rank of A?

SOLUTION: The rank of A is 2.

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

Symbolically,

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

(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?

SOLUTION: The matrix B would have orthonormal columns, so $B^T B = I_{2\times 2}$, while BB^T is the projection matrix for the column space of B.

26. In computing the best basis, what did we mean by "best"? (Be as specific as you can).

SOLUTION: "Best" can be described in two ways- Either

- The best one dimensional subspace is the subspace that maximizes the variance of the projection. The next one dimensional subspace is found the same way, and so on.
- The best $k-$dimensional basis is the one that minimizes the sum of the squares of the magnitude of the reconstruction error.

27. Suppose we have p points in $\mathbb{R}^n$,

$$\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_p\}$$

and we project each point to some (fixed) unit vector $\mathbf{u}$. Show that the mean of the scalar projections is the same as the projection of the mean vector.

SOLUTION: The scalar projections are:

$$\left\{\boldsymbol{u}^T\boldsymbol{x}_1, \boldsymbol{u}^T\boldsymbol{x}_2, \ldots, \boldsymbol{u}^T\boldsymbol{x}_p\right\}$$

so the mean of this data is:

$$\frac{1}{p}\boldsymbol{u}^T\boldsymbol{x}_i = \boldsymbol{u}^T\frac{1}{p}\sum_{i=1}^{p}\boldsymbol{x}_i = \boldsymbol{u}^T\bar{\boldsymbol{x}}$$

28. Continuing with the last problem, if we assume the mean of the data is zero, then show that the variance of the scalar projections to the unit vector $\mathbf{u}$ is given by:

$$\mathbf{u}^T\left(\frac{1}{p-1}\sum_{i=1}^{p}\mathbf{x}_i\mathbf{x}_i^T\right)\mathbf{u}$$

Hint: Recall that with our assumption of the mean, you can write the sample variance as $\sum_{i=1}^{p}(\mathbf{u}^T\mathbf{x}_i)^2$

SOLUTION: With this hint, it's pretty straightforward if you see this:

$$(\boldsymbol{u}^T\boldsymbol{x}_i)^2 = \boldsymbol{u}^T\boldsymbol{x}_i\boldsymbol{u}^T\boldsymbol{x}_i = \boldsymbol{u}^T\boldsymbol{x}_i\boldsymbol{x}_i^T\boldsymbol{u} = \boldsymbol{u}^T(\boldsymbol{x}_i\boldsymbol{x}_i^T)\boldsymbol{u}$$

Therefore, when you sum these together, you can factor $\boldsymbol{u}^T$ out in front and $\boldsymbol{u}$ on the right.

29. Suppose $\lambda, \mathbf{v}$ are the eigenvalue and (unit) eigenvector of the symmetric matrix C. Simplify the expression: $\mathbf{v}^T C\mathbf{v}$

SOLUTION: $\boldsymbol{v}^T C\boldsymbol{v} = \lambda$

Continuing, if $\mathbf{x} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2$, where $\lambda_1, \mathbf{v}_1$ and $\lambda_2, \mathbf{v}_2$ are eigenvalue, unit eigenvector pairs for the symmetric matrix C, show that

$$\mathbf{x}^T C\mathbf{x} = \lambda_1 c_1^2 + \lambda_2 c_2^2$$

SOLUTION: Just expand the dot products, and the orthogonality of the vectors will zero out most terms.