Mathematics for Machine Learning - Chapter 3 Solutions

Edwin Fennell

Note - I'm not going to write out the questions here since they are very, very inefficiently posed and no way am I going to TeX all of that.

3.1 We can more straightforwardly represent $\langle x, y\rangle$ as

$$x^T \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix} y$$

- Bilinearity - we note that tensor multiplication commutes with scalar multiplication and distributes over tensor addition. Therefore our function is linear over both x and y.
- Symmetry - the output of our function $\langle ., .\rangle$ is a scalar and thus is equal to its own transpose, so $\forall x, y$ we have

$$\langle x, y\rangle = x^T \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix} y = \left(x^T \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix} y \right)^T$$

$$= y^T \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}^T x = y^T \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix} x = \langle y, x\rangle$$

- Positive definite - for some reason this chapter doesn't cover diagonalisation so I guess we'll do this manually.
  Consider an arbitrary $v = (v_1, v_2)^T \in \mathbb{R}$. Then

$$\langle v, v\rangle = v_1^2 - 2v_1v_2 + 2v_2^2 = (v_1 - v_2)^2 + v_2^2 \geq 0$$

  with equality iff $v_1 = v_2 = 0$.

3.2 This is not an inner product since the matrix corresponding to this bilinear form is not symmetric. We observe that

$$\langle \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix} \rangle = 1$$

but also that

$$\langle \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \rangle = 0$$

This bilinear form is not symmetric and therefore is not an inner product

3.3 Before we start we should note that

$$x - y = \begin{pmatrix} 2 \\ 3 \\ 3 \end{pmatrix}$$

a. w.r.t. this inner product we have

$$\langle x - y, x - y \rangle = \begin{pmatrix} 2 & 3 & 3 \end{pmatrix} \begin{pmatrix} 2 \\ 3 \\ 3 \end{pmatrix} = 22$$

and so

$$||x - y|| = \sqrt{22}$$

b. w.r.t. this inner product we have

$$\langle x - y, x - y \rangle = \begin{pmatrix} 2 & 3 & 3 \end{pmatrix} \begin{pmatrix} 2 & 1 & 0 \\ 1 & 3 & -1 \\ 0 & -1 & 2 \end{pmatrix} \begin{pmatrix} 2 \\ 3 \\ 3 \end{pmatrix} = 47$$

and so

$$||x - y|| = \sqrt{47}$$

3.4 We define the "angle" θ w.r.t an inner product $\langle ., . \rangle$ (and its induced norm $||.||$) between two vectors x, y as

$$\theta = cos^{-1}\left(\frac{\langle x, y \rangle}{||x||, ||y||}\right)$$

where we define cos^{-1} to take values in $[0, \pi]$

a.

$$\langle x, y \rangle = x^T y = -3$$
$$||x|| = \sqrt{x^T x} = \sqrt{5}$$
$$||y|| = \sqrt{y^T y} = \sqrt{2}$$

and therefore we have

$$\theta = cos^{-1}\left(\frac{\langle x, y \rangle}{||x||, ||y||}\right) = cos^{-1}\left(\frac{-3}{\sqrt{10}}\right) \approx 161.6°$$

b.

$$\langle x, y \rangle = x^T A y = -11$$
$$||x|| = \sqrt{x^T A x} = \sqrt{18}$$
$$||y|| = \sqrt{y^T A y} = \sqrt{7}$$

and therefore we have

$$\theta = cos^{-1}\left(\frac{\langle x, y \rangle}{||x||, ||y||}\right) = cos^{-1}\left(\frac{-11}{\sqrt{126}}\right) \approx 168.5°$$

3.5 The way we proceed here is to use the Gram-Schmidt process to create an orthonormal basis for U. First we need a basis for U, so we had better construct one from our (not necessarily linearly independent) spanning set.

We do this via Gaussian elimination, as before. We start with a matrix where the rows are our spanning set:

$$\begin{pmatrix} 0 & -1 & 2 & 0 & 2 \\ 1 & -3 & 1 & -1 & 2 \\ -3 & 4 & 1 & 2 & 1 \\ -1 & -3 & 5 & 0 & 7 \end{pmatrix}$$

We switch the first and second rows to get

$$\begin{pmatrix} 1 & -3 & 1 & -1 & 2 \\ 0 & -1 & 2 & 0 & 2 \\ -3 & 4 & 1 & 2 & 1 \\ -1 & -3 & 5 & 0 & 7 \end{pmatrix}$$

We now reduce column 1 using row 1 to get

$$\begin{pmatrix} 1 & -3 & 1 & -1 & 2 \\ 0 & -1 & 2 & 0 & 2 \\ 0 & -5 & 4 & -1 & 7 \\ 0 & -6 & 6 & -1 & 9 \end{pmatrix}$$

We now reduce column 2 using row 2 to get

$$\begin{pmatrix} 1 & -3 & 1 & -1 & 2 \\ 0 & -1 & 2 & 0 & 2 \\ 0 & 0 & -6 & -1 & -3 \\ 0 & 0 & -6 & -1 & -3 \end{pmatrix}$$

The last two rows are exact clones so the last row just reduces to 0:

$$\begin{pmatrix} 1 & -3 & 1 & -1 & 2 \\ 0 & -1 & 2 & 0 & 2 \\ 0 & 0 & -6 & -1 & -3 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Therefore we have

$$\left\{ \begin{pmatrix} 1 \\ -3 \\ 1 \\ -1 \\ 2 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \\ 2 \\ 0 \\ 2 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ -6 \\ -1 \\ -3 \end{pmatrix} \right\}$$

as a basis for U. We denote these vectors as b_1, b_2, b_3 respectively.

We have a basis, now we can use the Gram-Schmidt process to orthonormalise it. Our first non-normalised basis vector is

$$u_1 = b_1 = \begin{pmatrix} 1 \\ -3 \\ 1 \\ -1 \\ 2 \end{pmatrix}$$

and so our first normalised basis vector is

$$e_1 = \frac{u_1}{||u_1||} = \frac{1}{4} \begin{pmatrix} 1 \\ -3 \\ 1 \\ -1 \\ 2 \end{pmatrix}$$

The next non-normalised basis vector is

$$u_2 = b_2 - \langle e_1, b_2 \rangle e_1 = \begin{pmatrix} 0 \\ -1 \\ 2 \\ 0 \\ 2 \end{pmatrix} - \frac{9}{16} \begin{pmatrix} 1 \\ -3 \\ 1 \\ -1 \\ 2 \end{pmatrix} = \frac{1}{16} \begin{pmatrix} -9 \\ 11 \\ 23 \\ 9 \\ 14 \end{pmatrix}$$

and so our second normalised basis vector is

$$e_2 = \frac{1}{12\sqrt{7}} \begin{pmatrix} -9 \\ 11 \\ 23 \\ 9 \\ 14 \end{pmatrix}$$

The final non-normalised basis vector is

$$u_3 = b_3 - \langle e_1, b_3 \rangle e_1 - \langle e_2, b_3 \rangle e_2$$

$$= \begin{pmatrix} 0 \\ 0 \\ -6 \\ -1 \\ -3 \end{pmatrix} - \frac{11}{16} \begin{pmatrix} 1 \\ -3 \\ 1 \\ -1 \\ 2 \end{pmatrix} - \frac{3}{16} \begin{pmatrix} -9 \\ 11 \\ 23 \\ 9 \\ 14 \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \\ -1 \\ 0 \\ 1 \end{pmatrix}$$

and so we have the final normalised basis vector as

$$e_3 = \frac{1}{\sqrt{3}} \begin{pmatrix} -1 \\ 0 \\ -1 \\ 0 \\ 1 \end{pmatrix}$$

a. Finally, we can now actually attack the question. We can now write the orthogonal projection of x onto U as

$$\sum_{i=1}^{3} \langle x, e_i \rangle e_i = \frac{23}{16} \begin{pmatrix} 1 \\ -3 \\ 1 \\ -1 \\ 2 \end{pmatrix} - \frac{1}{16} \begin{pmatrix} -9 \\ 11 \\ 23 \\ 9 \\ 14 \end{pmatrix} + \begin{pmatrix} -1 \\ 0 \\ -1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ -5 \\ -1 \\ -2 \\ 3 \end{pmatrix}$$

b. The distance between x and U is equal to the distance between x and its orthogonal projection onto U. This distance is therefore

$$\left\| \begin{pmatrix} -1 \\ -9 \\ -1 \\ 4 \\ 1 \end{pmatrix} - \begin{pmatrix} 1 \\ -5 \\ -1 \\ -2 \\ 3 \end{pmatrix} \right\| = \left\| \begin{pmatrix} -2 \\ -4 \\ 0 \\ 6 \\ -2 \end{pmatrix} \right\| = 2\sqrt{15}$$