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 What actually is the orthogonal projection? In this book it is motivated by "closest point" within a subset. Let's look what this means.

In a general infinite-dimensional vector space equipped with some arbitrary inner product, there need not exist a "closest point". So we restrict our problem to working with vector spaces equipped with n inner product that forces the vector space to be complete w.r.t the induced norm (known as a Hilbert space). We now proceed as follows:

First pick an arbitrary vector $u \in U$.