VECTOR SPACE PROBLEM SHEET

This paper asks you to prove some well known results. Although the algebra is easy the proofs are not entirely straightforward. There are marks assigned to the readability of the solution and also how well laid out and explained the steps you make are. (A good proof needs to be easy to follow: you need not comment on trivial algebra, but there should not be steps that are difficult to follow).

This looks very mathematical, but it helps to develop the tools and language that is used to describe machine learning.

1 An inner product $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ between vectors in a vector space $\mathcal{V}$ satisfies the following properties

(a) $\langle \boldsymbol{x}, \boldsymbol{x} \rangle \geq 0$ for all $\boldsymbol{x} \in \mathcal{V}$
(b) $\langle \boldsymbol{x}, \boldsymbol{x} \rangle = 0$ if and only if $\boldsymbol{x} = \mathbf{0}$
(c) $\langle \alpha \boldsymbol{x}, \boldsymbol{y} \rangle = \alpha \langle \boldsymbol{x}, \boldsymbol{y} \rangle$
(d) $\langle \boldsymbol{x}, \boldsymbol{y} + \boldsymbol{z} \rangle = \langle \boldsymbol{x}, \boldsymbol{y} \rangle + \langle \boldsymbol{x}, \boldsymbol{z} \rangle$
(e) $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \langle \boldsymbol{y}, \boldsymbol{x} \rangle$

The question explores properties of inner products.

(a) Consider the quadratic

$$q(t) = \langle \boldsymbol{x} + t\boldsymbol{y}, \boldsymbol{x} + t\boldsymbol{y} \rangle$$

By definition 1 of an inner product $q(t)$ must be non-negative and will only be 0 when $\boldsymbol{x} + t\boldsymbol{y} = 0$.

Expand $q(t)$ in the form $q(t) = At^2 + 2Bt + C$ to find A, B and C. For $q(t)$ to not change sign its roots (values of t such that $q(t) = 0$) must be complex (i.e. have an imaginary part), or possible have a double root. If there is a value of t such that $q(t) = 0$. Use the standard solutions to a quadratic of the form $q(t) = At^2 + 2Bt + C$ to show that for our $q(t)$ to never become negative then

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle^2 \leq \langle \boldsymbol{x}, \boldsymbol{x} \rangle \langle \boldsymbol{y}, \boldsymbol{y} \rangle .$$

This is the famous Cauchy-Schwarz inequality written in a very general form. [10 marks]

$$q(t) = \langle \boldsymbol{y}, \boldsymbol{y} \rangle t^2 + 2 \langle \boldsymbol{x}, \boldsymbol{y} \rangle t + \langle \boldsymbol{x}, \boldsymbol{x} \rangle$$

So that $A = \langle \boldsymbol{y}, \boldsymbol{y} \rangle$, $B = 2 \langle \boldsymbol{x}, \boldsymbol{y} \rangle$ and $C = \langle \boldsymbol{x}, \boldsymbol{x} \rangle$. The roots to the quadratic equation $q(t) = At^2 + Bt + C = 0$ are

$$t = \frac{-B \pm \sqrt{B^2 - 4AC}}{2A}.$$

The condition that these roots are complex is that $B^2 - 4AC < 0$, and where $B^2 - 4AC = 0$ indicating there is a double root. Thus $B^2/4 \leq AC$. Putting in the values of A, B and C then

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle^2 \leq \langle \boldsymbol{x}, \boldsymbol{x} \rangle \langle \boldsymbol{y}, \boldsymbol{y} \rangle .$$

(b) From an inner product we can define a norm $\|\boldsymbol{x}\| = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle}$. This clearly satisfies non-negativity and linearity. The only non-trivial property to show is that this norm satisfies the triangular inequality. Expand out $\|\boldsymbol{x} + \boldsymbol{y}\|$ and hence show that

$$\|\boldsymbol{x} + \boldsymbol{y}\|^2 = \|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2 + 2\langle \boldsymbol{x}, \boldsymbol{y} \rangle$$

Then use the Cauchy-Schwarz inequality to prove the triangular inequality. [10 marks]

We start form

$$\begin{aligned} \|\boldsymbol{x} + \boldsymbol{y}\|^2 &= \langle \boldsymbol{x} + \boldsymbol{y}, \boldsymbol{x} + \boldsymbol{y} \rangle \\ &= \langle \boldsymbol{x}, \boldsymbol{x} \rangle + \langle \boldsymbol{y}, \boldsymbol{y} \rangle + 2 \langle \boldsymbol{x}, \boldsymbol{y} \rangle \end{aligned}$$

Using the Cauchy-Schwarz inequality $|\langle \boldsymbol{x}, \boldsymbol{y} \rangle|^2 \leq \langle \boldsymbol{x}, \boldsymbol{x} \rangle \langle \boldsymbol{y}, \boldsymbol{y} \rangle$ so that

$$\begin{aligned} \|\boldsymbol{x} + \boldsymbol{y}\|^2 &\leq \langle \boldsymbol{x}, \boldsymbol{x} \rangle + \langle \boldsymbol{y}, \boldsymbol{y} \rangle + 2\sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle \langle \boldsymbol{y}, \boldsymbol{y} \rangle} \\ &= \|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2 + 2\|\boldsymbol{x}\| \|\boldsymbol{y}\| = (\|\boldsymbol{x}\| + \|\boldsymbol{y}\|)^2 \end{aligned}$$

Taking the square root (which is monotonic so does not change the inequality)

$$\|\boldsymbol{x}+\boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$$

Thus proving the triangular inequality for $\|\boldsymbol{x}\| = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle}$.

End of question 1

2 Random variables, X, Y, etc. form a vector space (i.e. they satisfy properties such as closure under addition and scalar multiplication). Furthermore we can define an inner product between random variables as

$$\langle X,Y \rangle = \mathbb{E}\left[XY\right]$$

(i.e. the expectation of the random variable $Z = XY$)

(a) Use the Cauchy-Schwarz inequality to show that

$$\mathrm{Cov}(X,Y)^2 \leq \mathrm{Var}(X)\,\mathrm{Var}(Y)$$

hence show that the Pearson correlation is between -1 and 1. [10 marks]

Let $\mu = \mathbb{E}\left[X\right]$ and $\nu = \mathbb{E}\left[Y\right]$ then

$$\mathrm{Cov}(X,Y) = \mathbb{E}\left[(X-\mu)(Y-\nu)\right]$$

From the Cauchy-Schwarz inequality

$$\begin{aligned}\mathrm{Cov}(X,Y)^2 &= \mathbb{E}\left[(X-\mu)(Y-\nu)\right]^2 = \langle X-\mu, Y-\mu \rangle^2 \\ &\leq \langle X-\mu, X-\mu \rangle \langle Y-\mu, Y-\mu \rangle = \mathbb{E}\left[(X-\mu)^2\right]\mathbb{E}\left[(Y-\nu)^2\right] = \mathrm{Var}(X)\,\mathrm{Var}(Y).\end{aligned}$$

which proves the inequality. The Pearson correlation is defined as

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}}$$

Thus $\rho(X,Y)^2 = \mathrm{Cov}(X,Y)^2/(\mathrm{Var}(X)\,\mathrm{Var}(Y))$, but we have already shown that $\mathrm{Cov}(X,Y)^2 \leq \mathrm{Var}(X)\,\mathrm{Var}(Y)$ so $\rho(X,Y)^2 \leq 1$ and $\rho(X,Y)$ will lie in the interval $[-1,1]$.

(b) Show that for vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ that the inner product

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \sum_{i=1}^{n} w_i x_i y_i$$

satisfies the condition of an inner product provided $w_i > 0$ for all i. Write down the norm and distance induced by this inner product and provide an interpretation of what this distance means.

[10 marks]

We consider the five condition on the inner product

(i) Non-negativity follows from

$$\langle \boldsymbol{x}, \boldsymbol{x} \rangle = \sum_{i=1}^{n} w_i x_i^2 \geq 0$$

since for each term in the sum $w_i x_i^2 \geq 0$.

(ii) For $\langle \boldsymbol{x}, \boldsymbol{x} \rangle = 0$ each term in the sum must be zero. Since $w_i > 0$ this implies $x_i = 0$ for all i, that is, $\boldsymbol{x} = \boldsymbol{0}$.

(iii) Linearity follows trivially since

$$\begin{aligned} \langle a\boldsymbol{x}, \boldsymbol{y} \rangle &= \sum_{i=1}^{n} w_i (a x_i) y_i \\ &= a \sum_{i=1}^{n} w_i x_i y_i = a \langle \boldsymbol{x}, \boldsymbol{y} \rangle \end{aligned}$$

(iv) The distributive law again follows trivially

$$\begin{aligned} \langle \boldsymbol{x}, \boldsymbol{y} + \boldsymbol{z} \rangle &= \sum_{i=1}^{n} w_i x_i (y_i + z_i) \\ &= \sum_{i=1}^{n} w_i x_i y_i + \sum_{i=1}^{n} w_i x_i z_i \\ &= \langle \boldsymbol{x}, \boldsymbol{y} \rangle + \langle \boldsymbol{x}, \boldsymbol{z} \rangle \end{aligned}$$

(v) Finally symmetry follows from the commutativity of multiplication

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \sum_{i=1}^{n} w_i x_i y_i = \sum_{i=1}^{n} w_i y_i x_i = \langle \boldsymbol{y}, \boldsymbol{x} \rangle$$

The induced norm for this inner product is

$$\|x\| = \sqrt{\sum_{i=1}^{n} w_i x_i^2}$$

while the induced distance is

$$d(\boldsymbol{x}, \boldsymbol{y}) = \sqrt{\sum_{i=1}^{n} w_i (x_i - y_i)^2}$$

A natural interpretation of this is that this is a generalisation of the Euclidean distance where we rescale each axis by a factor w_i and then compute the normal Euclidean distance.

End of question 2

END OF PAPER