Name: Supritha Konaje Student ID: 32864477

PROBLEM SHEET 1 FOR ADVANCED MACHINE LEARNING (COMP6208)

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 $\boldsymbol{x} = \boldsymbol{0}$
(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]

Using the axiom, $\langle x, y+z \rangle = \langle x, y \rangle + \langle x, z \rangle$ and $\langle \alpha x, y \rangle = \alpha \langle x, y \rangle$

$\therefore q(t) = \langle x, x \rangle + \langle x + ty \rangle + \langle ty + x \rangle + \langle ty, ty \rangle$

Using the axiom, $\langle x, y \rangle = \langle y, x \rangle$ and $\langle \alpha x, y \rangle = \alpha \langle x, y \rangle$

$\therefore q(t) = \langle x, x \rangle + 2t \langle x, y \rangle + t^2 \langle y, y \rangle$ — (1)

Comparing with $At^2 + 2Bt + C$

$A = \langle y, y \rangle$; $B = \langle x, y \rangle$; $C = \langle x, x \rangle$

Using the axiom, $\langle x, x \rangle \geq 0$ for all $x \in V$ — (2)

To minimize, we get the derivative of $q(t)$

$q'(t) = 2At + 2B = 0 \quad \therefore t = -\frac{B}{A} \Rightarrow \boxed{t = -\frac{\langle x, y \rangle}{\langle y, y \rangle}}$

Substitute t in (1)

$\langle x, x \rangle + 2\left(-\frac{\langle x, y \rangle}{\langle y, y \rangle}\right) \langle x, y \rangle + \left(-\frac{\langle x, y \rangle}{\langle y, y \rangle}\right)^2 \langle y, y \rangle$

$\langle x, x \rangle - 2\frac{\langle x, y \rangle^2}{\langle y, y \rangle} + \frac{\langle x, y \rangle^2}{\langle y, y \rangle} = \langle x, x \rangle - \frac{\langle x, y \rangle^2}{\langle y, y \rangle}$

From (2), $\langle x, x \rangle - \frac{\langle x, y \rangle^2}{\langle y, y \rangle} \geq 0 \Rightarrow \langle x, x \rangle \geq \frac{\langle x, y \rangle^2}{\langle y, y \rangle}$

$\therefore \langle x, y \rangle^2 \leq \langle x, x \rangle \langle y, y \rangle \quad \therefore$ Proved

10

(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

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

Then use the Cauchy-Schwarz inequality to prove the triangular inequality.

[10 marks]

Given $\|x\| = \sqrt{<x,x>}$ — 1

Consider $\|x+y\|^2 = <x+y, x+y>$

$= <x,x>^2 + 2|<x,y>| + <y,y>^2$

From 1

$\|x+y\|^2 = \|x\|^2 + 2|<x,y>| + \|y\|^2$ — 2

Using Cauchy Schwarz inequality,

i.e. $<x,y>^2 \leq <x,x><y,y>$

we get,

$<x,y> \leq \sqrt{<x,x><y,y>}$ — 3

From 2 & 3, we get

$\|x+y\|^2 = \|x\|^2 + 2|<x,y>| + \|y\|^2 \leq \|x\|^2 + 2\|x\|\|y\| + \|y\|^2$

$= (\|x\| + \|y\|)^2$

that is, $\|x+y\|^2 \leq (\|x\| + \|y\|^2)$

Taking square root, we get

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

$\therefore$ Proved

$\overline{10}$

End of question 1

(a) $\overline{10}$ (b) $\overline{10}$ Total $\overline{20}$

TURN OVER

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]

Given $<X,Y> = E[XY]$ — 1

Using Cauchy-Schwarz inequality: $|E(XY)|^2 \leq E(X^2)\,E(Y^2)$ 2

To prove covariance inequality; Let $\alpha = E(X)$ and $\beta = E(Y)$, then

$$|Cov(X,Y)|^2 = |E((X-\alpha)(Y-\beta))|^2$$

$$= |<X-\alpha, Y-\beta>|^2$$

$$= <X-\alpha, X-\alpha><Y-\beta, Y-\beta>$$

$$= E((X-\alpha)^2)\,E((Y-\beta)^2) \quad \text{(From 1)}$$

$\therefore$ From 2, $|Cov(X,Y)|^2 \leq E((X-\alpha)^2)\,E((Y-\beta)^2)$

Variance is Expectation of the squared deviation of random no from sample mean

$$\therefore \boxed{|Cov(X,Y)^2| \leq Var(X)\,Var(Y)} \quad \text{— 3}$$

$$\therefore \frac{|Cov(X,Y)|^2}{Var(X)Var(Y)} \leq 1 \quad \text{— 4}$$

Pearson's coefficient: $\rho = \dfrac{|Cov(X,Y)|}{\sqrt{var(X)\,Var(Y)}}$ (Squaring)

$\rho^2 = \dfrac{|Cov(X,Y)|^2}{Var(X)\,Var(Y)}$ $\Rightarrow$ From 4; $\rho^2 \leq 1$ (taking square root)

$\Rightarrow \rho \leq \pm 1 \Rightarrow \boxed{-1 \leq \rho \leq +1}$

$\therefore$ Proved that Pearson's coefficient lies between -1 & $+1$

10

(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]

Consider $< , >$ to be the inner product of $\mathbb{R}^n$.

The weight basis of $\mathbb{R}^n$ is $(w_1, w_2, \ldots, w_n)$ & the scalars are $(x_1, x_2, \ldots, x_n)$ and $(y_1, y_2, \ldots, y_n)$

$\therefore x = x_1 w_1 + x_2 w_2 + \ldots + x_n w_n$

$y = y_1 w_1 + y_2 w_2 + \ldots + y_n w_n$

$\therefore <x, y> = <x_1 w_1 + x_2 w_2 + \ldots + x_n w_n, y_1 w_1 + y_2 w_2 + \ldots + y_n w_n>$

$= < \sum_{i=1}^{n} x_i w_i, \sum_{i=1}^{n} y_i w_i >$

$(\because <x, y> = x \cdot y)$

$\boxed{<x, y> = \sum_{i=1}^{n} w_i x_i y_i}$

& it satisfies only if $w_i > 0$ and w_i is common for both

Norm:

$\because w_i$ is all positive numbers

The weighted norm is given by $\|x\| = \sqrt{<x, x>}$

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

Distance: Weighted distance is given by

$dist(x, y) = \|x - y\| = \sqrt{(x_1 w_1 - y_1 w_1)^2 + (x_2 w_2 - y_2 w_2)^2 + \ldots (x_n w_n - y_n w_n)^2}$

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

$\frac{\quad}{10}$

End of question 2

(a) $\frac{\quad}{10}$ (b) $\frac{\quad}{10}$ Total $\frac{\quad}{20}$

