Name: Han Zhang Student ID: 32672446

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]

As mentioned, $q(t) = \langle x + ty, x + ty \rangle \geqslant 0$ by a

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

$= \langle x, x \rangle + 2\langle x, ty \rangle + \langle ty, ty \rangle$

$= \langle x, x \rangle + 2t\langle x, y \rangle + t^2\langle y, y \rangle$ by C

Thus the quadratic equcation:

$t^2\langle y, y \rangle + 2\langle x, y \rangle t + \langle x, x \rangle \geqslant 0$

If $\langle y, y \rangle = 0$, $y = 0$. $\langle x, 0 \rangle = \langle x, 0 \rangle + \langle x, 0 \rangle$ by d

so $\langle x, 0 \rangle = 0$. the quadratic equcation $\langle x, x \rangle \geqslant 0$ is always true.

Otherwise $\langle y, y \rangle$ is bigger than 0. So $q(t)$ is a polynomial of degree 2. As the results, its discriminant Δ must be less or equal than 0.

$\Delta = (2\langle x, y \rangle)^2 - 4\langle x, x \rangle\langle y, y \rangle \leq 0$

So we get inequation: $\langle x, y \rangle^2 \leq \langle x, x \rangle \langle y, y \rangle$

$\overline{\quad}$ / 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]

As the definition of norm:

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

As mentioned in 1.a:

$$\langle x,y \rangle^2 \leq \langle x,x \rangle \langle y,y \rangle$$

$$\|\langle x,y \rangle\| \leq \sqrt{\langle x,x \rangle \langle y,y \rangle} = \|x\| \cdot \|y\|$$

So the equation: $\|x+y\|^2 = \|x\|^2 + \|y\|^2 + 2\langle x,y \rangle$

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

In the end, we prove that

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

So the triangular inequality can be proved:

$$\|x+y\| \leq \|x\| + \|y\|,$$

since both sides are nonnegative.

$\overline{}$ / 10

End of question 1

(a) ___ / 10 (b) ___ / 10 Total ___ / 20

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]

To prove the covariance inequality, Let $\mu = E(X)$, $V = E[Y]$,

as we know: $cov(X,Y) = E[(X-\mu)(Y-V)]^2$

$Var(X) = E[(X-\mu)^2]$

$Var(Y) = E[(Y-V)^2]$

$cov(X,Y)^2 = E[(X-\mu)(Y-V)]^2$

$= ||\langle X-\mu, \langle Y-V\rangle||^2$

by cauchy-schwarz inequality:

$||\langle X-\mu, \langle Y-V\rangle||^2 \leq \langle X-\mu, X-\mu\rangle \langle Y-V, Y-V\rangle$

$= E[(X-\mu)^2]\, E[(Y-V)]^2$

$= Var(X)\, Var(Y)$

So $cov(X,Y)^2 \leq var(X)\, Var(Y)$, we get

$$-1 \leq \frac{cov(X,Y)}{\sqrt{var(X)\, var(Y)}} \leq 1$$

So the pearson correlation $\rho(X,Y) = \dfrac{cov(X,Y)}{\sqrt{var(X)\, var(Y)}}$ is between -1 and 1.

$\overline{}$
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]

Let $x = [x_1, x_2 \cdots x_n]^T$, $y = [y_1, y_2 \cdots y_n]^T$

If $\langle x, y \rangle = \sum_{i=1}^{n} w_i x_i y_i$ satisfies the condition of inner product, then

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

as $x_i^2 \geq 0$ and x_i can not be always 0, so $w_i > 0$.

$$\langle x, y \rangle = [x_1, x_2 \cdots x_n] \begin{bmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & \cdots & w_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = x^T \Sigma y = \sum_{i=1}^{n} x_i w_i y_i$$

where $\Sigma = \begin{bmatrix} w_1 & \cdots & 0 \\ \vdots & w_2 & \vdots \\ \vdots & & \ddots \\ 0 & \cdots & w_n \end{bmatrix}$ $\Sigma \in R^n$

norm: $\|x\| = \sqrt{\langle x, x \rangle} = \sqrt{x^T \Sigma x} = \Sigma^{\frac{1}{2}} x$

distance: $\|x - y\| = \sqrt{\langle x-y, x-y \rangle} = \sqrt{(x-y)^T \Sigma (x-y)}$

The distance is known as mahalanobis distance. when Σ is identity matrix, It becomes Euclidean distance. It's a effective method to measure a distance when the variables (x, y) have some correlation since It can eliminate the correlation between variables (x, y).

$\overline{10}$

End of question 2

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