

Homework 2

September 5, 2019

Cameron Taylor

Mathematical Foundations of ML

Homework 2

0.0.1 Problem 1

In the first lecture, we talked about norms and inner products, specifically covering the properties of these two operations and discussing some of their uses. For norms, we discussed the different versions such as l_2 norm, l_1 norm, l_∞ , and l_p norm. We also discussed different domains where this operation can be defined such as R^n , infinite sequences, real-valued functions, etc. For inner products we discussed its relationship to orthogonality and some standard and slightly less standard examples of spaces where it can be defined. We also talked about how a valid inner product space will have an induced norm in the form of $\|x\| = \sqrt{\langle x, x \rangle}$ and some other results that can be derived from the norm such as Cauchy-Schwarz ($|\langle x, y \rangle| \leq \|x\| \|y\|$), the parallelogram law ($\|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + 2\|y\|^2$), and the polarization identity ($\frac{\|x+y\|^2 - \|x-y\|^2}{4}$). In addition we covered how these inner products relate to the concept of angle between vectors and how that might be useful.

In the second lecture, we discussed linear approximation in a Hilbert space, where a Hilbert space is a space with an inner product and induced norm. The goal of the lecture was to show the best way to find the closest point $\hat{x} \in T$, which is a subspace of S , to some other point $x \in S$. This involves finding some α 's to multiply with a basis of T such that $x - \hat{x}$ is orthogonal to the subspace T . We then talked about how expanding out the solution for this problem provides us with a Gram matrix and discussed the useful properties of this type of matrix such as being always invertible and conjugate symmetric. Finally, we finished with an example of how to use this approach and approximate e^t .

0.0.2 Problem 2

Prove $||x| - |y|| \leq \|x - y\|$ is true in normed linear space.

$$\|x\| = \|(x - y) + y\| \leq \|x - y\| + \|y\|$$

This statement is true by the triangle inequality. From here if we replace the left side with $||x| - |y||$ we get...

$$||x| - |y|| \leq \|x - y\| + \|y\|$$

Now we can subtract $\|y\|$ from both sides to get...

$$\|x\| - \|y\| \leq \|x - y\|$$

We can say that $\|x\| - \|y\|$ will always be bounded by the length of the vector formed by subtracting y from x .

0.0.3 Problem 3

a) For B_1 $r = 7$ because $\|x\|_{l_1}$ is 7 when $x = [-3 \ 4]$.

For B_2 $r = 5$ because $\|x\|_{l_2}$ is 5 when $x = [-3 \ 4]$.

For B_∞ $r = 4$ because $\|x\|_{l_\infty}$ is 4 when x is $[-3 \ 4]$.

b) B_b breaks the homogeneity principle if you take $x = [1 \ -2]$. We then multiply x by -1 giving us $[-1 \ 1]$, which should have the same value as x since $\|ax\| = |a| * \|x\|$. However, looking at the unit ball for B_b we have that $[-1 \ 1]$ is outside the boundary meaning that its value is greater than 1. This should be impossible for a unit ball that is created from a valid norm.

B_c breaks the triangle inequality principle if you take $x = [1 \ 1]$ and $y = [-1 \ 1]$. $\|x + y\| = \|[0 \ 1]\|_2$. $\|x\| = 1$ and $\|y\| = 1$ and $\|x + y\| = \|[0, 2]\|$. Let the vector $[0, \alpha]$ be the vector on the boundary of the unit ball. Since $[0, \alpha]$ is colinear with $[0, 2]$ and we know that it must have twice the norm of $[0, 1]$ by the property of homogeneity. Therefore, we can say that $\|[0, 2]\|$ must be $> 2 * \|[0, \alpha]\|$. $\|[0, \alpha]\|$ is 1 so that means $\|x + y\| > 2$, but also ≤ 2 by the triangle inequality so B_c breaks the triangle inequality making it an invalid norm.

c) For all three matrices A below $\|Ax\|$ represents a valid norm when A is non singular (i.e. $\text{Det}(A) \neq 0$). This is because Ax is still a vector so The triangle inequality and homogeneity principle still hold for all x . However it is possible that if A was singular it could map multiple x to zero breaking the property that $\|x\| \geq 0$ and $\|x\| = 0$ iff $x = 0$.

$B_a = \|Ax\|_\infty$, where $A = \begin{bmatrix} 1/3 & 0 \\ 0 & 1 \end{bmatrix}$. Since the $\text{Det}(A) = 1/3 \neq 0$, we know that the null space of A is 0 which means B_a is a valid norm.

$B_d = \|Ax\|_1$, where $A = \begin{bmatrix} 1 & 0 \\ 0 & 1/2 \end{bmatrix}$. Since the $\text{Det}(A) = 1/2 \neq 0$, we know that the null space of A is 0 which means B_d is a valid norm.

$B_e = \|Ax\|_1$, where $A = \begin{bmatrix} 1/2 & -1/2 \\ 1/4 & 1/4 \end{bmatrix}$. This matrix comes from a composition of stretching matrix S , similar to the two above, and a rotation matrix R . The image is rotated clockwise 45 degrees after being stretched giving the translation matrix $T = SR$. Since the $\text{Det}(A) = \frac{1}{4} \neq 0$, we know that the null space of A is 0 which means B_e is a valid norm.

0.0.4 Problem 4

a) $|\langle x, y \rangle| \leq \|x\|_\infty \|y\|_1$

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

$$\|x\|_\infty \|y\|_1 = \max(x) \sum_{i=1}^n y_i = \sum_{i=1}^n \max(x) (y_i)$$

By definition $\max(x)$ will be \geq all elements of x so for each term in the sum the value of $\max(x)y_i$ will be $\geq x_i y_i$ which implies that the sum of the terms will also be greater than or equal to.

b) Let $y = [y_1, y_2, \dots, y_n]$ where $y_i = \text{sign}(x_i) * 1$. From this we can say that $\|y\|_2 = \sqrt{N}$ where N is the length of y . Using Cauchy-Schwartz we can replace $\|y\|_2$ with \sqrt{N} giving us $\langle x, y \rangle \leq \sqrt{N} \|x\|_2$. Since $\text{sign}(x_i) * x_i = |x_i|$ we can also say that $\langle x, y \rangle = \|x\|_1$. This gives us $\|x\|_1 \leq \sqrt{N} \|x\|_2$, proving the inequality.

c) Since we know that the angle between x and y must be zero in order to get the maximum value from the inner product, we know that x must be equal to y scaled by some constant. We also know that we want the l_2 norm of x to be equal to 1 so we must divide y by the l_2 norm in order to scale it to the size we want.

$$\langle x, y \rangle = \sum_{i=1}^n x_i y_i \text{ where } x_i = \frac{y_i}{\|y\|_2}$$

d) We know that $\langle x, y \rangle = x_1 y_1 + x_2 y_2 + \dots + x_n y_n$. $\max(\text{sum}(\langle x, y \rangle)) = \text{sum}(\max(x_i y_i))$ so we know that by maximizing each individual term we will maximize the sum. The range of possible values for x_i is $[-1, 1]$ so in order to get the maximum contribution from each term $|x_i|$ should be 1. We also know that a sum of all positive values will always be greater than a sum of the same values but some being negative so we need each term to be positive.

$$\langle x, y \rangle = \sum_{i=1}^n x_i y_i \text{ where } x_i = \text{sign}(y) * 1$$

e) We know that is equivalent to a weighted sum where the sum of the absolute value of the weights must be ≤ 1 . For a weighted sum $x_1 y_1 + x_2 y_2 + \dots + x_n y_n$ we can find the maximum by weighting the term containing $\text{argmax}(y)$ as 1 (or -1 depending on the sign of y) and all others zero.

This can be shown by assuming that some other weighting would achieve the maximum so instead of $\text{sign}(y) * \text{argmax}(y) * 1$ we have $\text{sign}(y) * \text{argmax}(y) * (1 - \alpha) + \alpha * y_i$ where i is not the index of $\text{argmax}(y)$. For any value of α we are reducing the sum by $|\text{argmax}(y) * \alpha|$ and increasing the sum by $|\alpha * y_i|$ for some i other than index of $\text{argmax}(y)$. Obviously $|\text{argmax}(y) * \alpha| \geq |\alpha * y_i|$ for all i other than index of $\text{argmax}(y)$. So no matter which y_i or multiple y_i 's we multiply portions of α with we will always reduce the sum by more than we add.

$$\langle x, y \rangle = \sum_{i=1}^n x_i y_i \text{ where } x_i = \text{sign}(y) * 1 \text{ if } y_i = \text{argmax}(y); \text{ otherwise } 0.$$

0.0.5 Problem 5

a) By taking the formula for $\|x\|_{l_2}$ and applying A to x we get...

$$\begin{aligned} & \sqrt{(3x_1 + 3x_2)^2 + (-\frac{1}{2}x_2 + \frac{1}{2}x_2)^2} \\ &= \frac{37}{4}x_1^2 + \frac{35}{2}x_1x_2 + \frac{37}{4}x_2^2 \leq 1 \end{aligned}$$

This represents the equation of our unit ball which is plotted in python below.

```
[96]: import numpy as np
import matplotlib.pyplot as plt
import pylab
import math
```

```
[116]: def plot_lA_norm():
    xs_temp = [x for x in 2.0*np.random.randn(100000,2)]
    xs = []
    ys = []

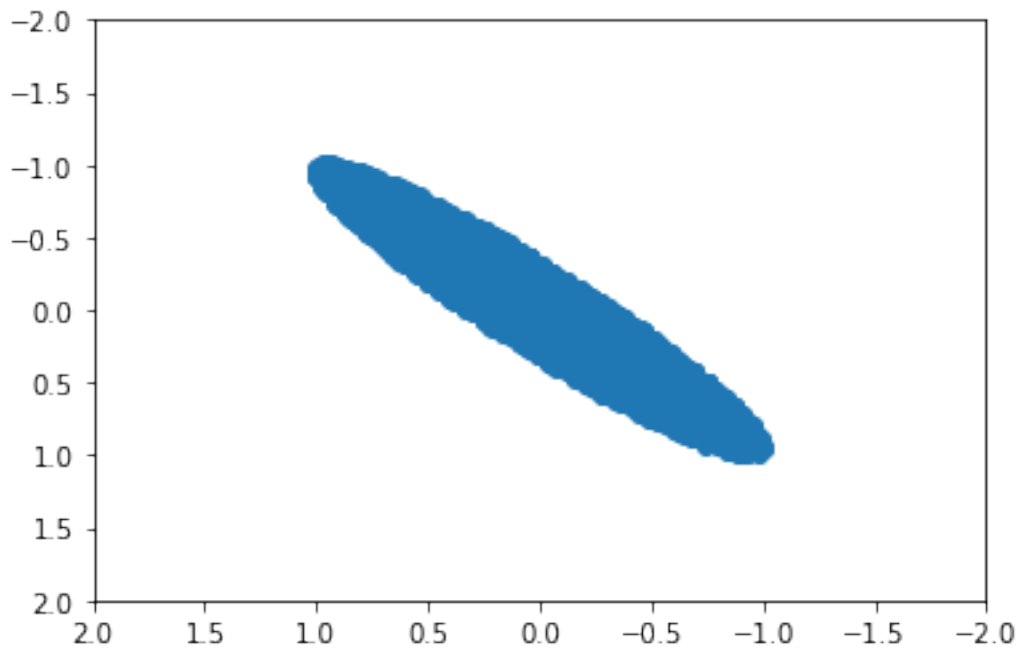
    angle = math.pi / 4
    val = math.cos(angle)

    for x in xs_temp:
        if (37/4)*x[0]**2 + (35/2)*x[0]*x[1] + (37/4)*x[1]**2 < 1:
            xs.append(x.item(0))
            ys.append(x.item(1))

    pylab.axis([2, -2, 2, -2])

    plt.scatter(xs, ys)
    plt.show()

plot_lA_norm()
```



b) Going through the properties of a norm we can see that when A has a determinant of zero, there may be vectors other than the zero vector which would allow $\|Ax\| = 0$ when $x \neq 0$. Therefore, only matrices A with non-zero determinants will create valid norms.

0.0.6 Problem 6

a) Assume $\langle x, y \rangle_Q = y^T Q x$ for matrices Q with a negative or zero on the diagonal represents a valid inner product in R^n . This means that $\langle x, x \rangle_Q \geq 0$ for all $x \neq 0$.

If we let $x = [0, 0, \dots, 1, 0, \dots, 0]$ such that the 1 is in the i th position where i is the row of Q where the negative value exists then...

$$\begin{aligned} x^t Q x &\text{ will be equal to } [Q_{i,1}, Q_{i,2}, \dots, Q_{i,i}, \dots, Q_{i,n}]x. \\ &= Q_{i,i} \end{aligned}$$

We know that $\langle x, x \rangle_Q \geq 0$ for all $x \neq 0$ and we also know that $Q_{i,i}$ is the negative value or zero on the diagonal. If it is zero we have $= 0$ for $x \neq 0$, otherwise we have $\langle x, x \rangle_Q < 0$. Therefore we have a contradiction and our assumption that $\langle x, y \rangle_Q = y^T Q x$ is a valid inner product for matrices Q with negative values on the diagonal is not true.

b) Assume $\langle x, y \rangle_Q = y^T Q x$ for non symmetric matrices Q represents a valid inner product in R^n . This means that $\langle x, y \rangle_Q$ must be equal to $\langle y, x \rangle_Q$.

$$\begin{aligned} \langle x, y \rangle_Q &= \sum_{i=1}^n x_i \sum_{j=1}^n y_j Q_{ij} \\ \langle y, x \rangle_Q &= \sum_{i=1}^n y_i \sum_{j=1}^n x_j Q_{ij} \\ \sum_{i=1}^n x_i \sum_{j=1}^n y_j Q_{ij} &= \sum_{i=1}^n y_i \sum_{j=1}^n x_j Q_{ij} \end{aligned}$$

Expanding the sums we can see that for each term which contains the product of $x_i y_j$ in the first sum will be multiplied by $Q_{i,j}$ while the other side will be multiplied by $Q_{j,i}$. Therefore this equation only holds true for matrices Q such that $Q_{i,j} = Q_{j,i}$. This means that Q must be symmetric giving us a contradiction implying that $\langle x, y \rangle_Q = y^T Q x$ is only a valid inner product in R^n for symmetric matrices Q.

c) Assume $\langle x, y \rangle_Q = y^T Q x$, where Q is not a symmetric positive definite matrix, is a valid inner product in R^n . This means that Q must not break any of the axioms of a valid inner product.

Axiom 1 states that $\langle x, y \rangle = \langle y, x \rangle$ and we know that this is only true if Q is symmetric so Q must be symmetric.

Axiom 3 states that $\langle x, x \rangle \geq 0$ and $\langle x, x \rangle = 0$ iff $x = 0$. $\langle x, x \rangle_Q = x^t Q x$ and we know that the definition of positive definite matrices is that $x^t Q x > 0, \forall x \in R^n, x \neq 0$ so for axiom 3 to hold Q must be positive definite.

Now we must show that $\langle x, y \rangle_Q = y^T Q x$ is a valid inner product since we know that if it is valid Q must be symmetric positive definite. Assuming Q is symmetric positive definite we know that axioms 1 and 3 hold, so we must show that axiom 2 also holds.

We want to show that $\langle ax + by, z \rangle \neq a \langle x, z \rangle + b \langle y, z \rangle$ when Q is symmetric positive definite.

$$\langle ax + by, z \rangle = z^T Q(ax + by)$$

$$a \langle x, z \rangle + b \langle y, z \rangle = az^T Qx + bz^T Qy$$

Since a and b are scalars the order of their multiplication does not matter.

$$= z^T Qax + z^T Qby$$

$$\text{Let } Z' = z^T Q, X' = ax \text{ and } Y' = by$$

$$= Z'X' + Z'Y'$$

By the distributive property of matrices we can factor Z' .

$$= Z'(X' + Y')$$

Substituting back in for Z', X', Y' we have...

$$= z^T Q(ax + by)$$

$$\langle ax + by, z \rangle = a \langle x, z \rangle + b \langle y, z \rangle$$

We have proven that Q must be symmetric positive definite in order for $\langle x, y \rangle_Q$ to be a valid inner product and we have also shown that if Q is symmetric positive definite all three axioms of a valid inner product hold so we can conclude that $\langle x, y \rangle_Q$ is a valid inner product *iff* Q is symmetric positive definite.

[]: