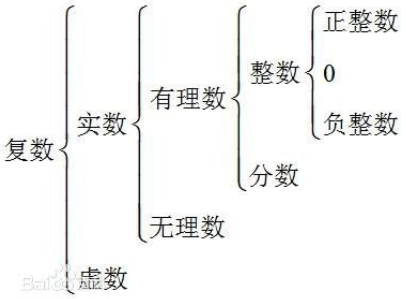
Contents

Chapter 1

Math

1.1 定义

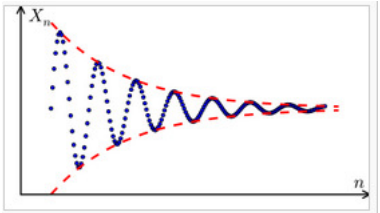


[1]

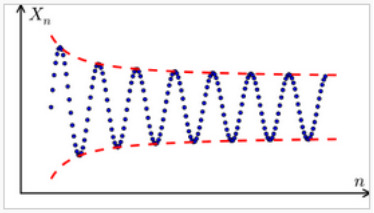
infimum——下确界，缩写 inf，复数 infima
supremum——上确界，缩写 sup，复数 suprema

1.1.1 柯西序列(Cauchy sequence)

[2] is a sequence whose elements become arbitrarily close to each other as the sequence progresses. More precisely, given any small positive distance, all but a finite number of elements of the sequence are less than that given distance from each other.



(a) The plot of a Cauchy sequence (x_n), shown in blue, as x_n versus n. If the space containing the sequence is complete, the "ultimate destination" of this sequence (that is, the limit) exists.



(b) A sequence that is not Cauchy. The elements of the sequence fail to get arbitrarily close to each other as the sequence progresses.

1.1.2 完备(Completeness)(by mg:柯西序列收敛与此)

[2] A metric space X in which every Cauchy sequence converges to an element of X is called **complete**.

1.1.3 紧空间(Compact spaces)(by mg:都有子序列收敛于此)

[2] A metric space M is **compact** if every sequence in M has a subsequence that converges to a point in M. This is known as sequential compactness and, in metric spaces (but not in general topological spaces), is equivalent to the topological notions of countable compactness and compactness defined via open covers.

1.1.4 赫米特矩阵 (Hermitian matrix)

共轭转置（conjugate transpose）等于自身的矩阵（主要针对复数情况）：

$$a_{ij} = \overline{a_{ji}} \text{ 或者 } A = \overline{A^T}$$

可以认为是对实对称矩阵的扩展。

1.1.5 格拉姆矩阵 (Gramian matrix)

In linear algebra, the Gram matrix (Gramian matrix or Gramian) of a set of vectors $v_1, \ldots, v_n$ in an inner product space is the Hermitian matrix of **inner products**, whose entries are given by $G_{ij} = \langle v_i, v_j \rangle$

1.1.6 正定

正定函数

正定矩阵 positive definite matrix

[3] In linear algebra, a symmetric $n \times n$ real matrix M is said to be **positive definite** if the scalar $z^T M z$ is positive for every non-zero column vector z of n real numbers.

例如，$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ 是正定的。Seen as a real matrix, it is symmetric, and, for any non-zero column vector z with real entries a and b, one has

$$z^T I z = [a\ b] \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = a^2 + b^2$$

1.2 拓扑

1.2.1 Topology(by mg:重点在连续)

[4]

topology is concerned with the properties of space that are preserved under **continuous deformations**, such as stretching and bending, but not tearing or gluing. This can be studied by considering a collection of subsets, called open sets, that satisfy certain properties, turning the given set into what is known as a topological space. Important topological properties include connectedness and compactness.

Formally, let X be a set and let τ be a family of subsets of X. Then τ is called a topology on X if:

1. Both the empty set and X are elements of τ
2. Any union of elements of τ is an element of τ
3. Any intersection of finitely many elements of τ is an element of τ

If τ is a topology on X, then the pair (X, τ) is called a **topological space**. The notation X_τ may be used to denote a set X endowed with the particular topology τ.

1.2.2 Topological vector space

[5]

A topological vector space X is a vector space over a topological field K (most often the real or complex numbers with their standard topologies) that is endowed with a topology such that vector addition $X \times X \to X$ and scalar multiplication $K \times X \to X$ are **continuous functions** (where the domains of these functions are endowed with product topologies).

1.3 度量

1.3.1 Metric space(by mg: 定义了距离)

[6]

A metric space is a set for which distances between all members of the set are defined. Those distances, taken together, are called a **metric** on the set.

正式定义：A metric space is an ordered pair (M, d) where M is a set and d is a metric on M, i.e., a function, $d : M \times M \to R$, such that for any $x, y, z \in M$, the following holds:

1.	$d(x, y) \geq 0$	non-negativity or separation axiom
2.	$d(x, y) = 0 \Leftrightarrow x = y$	Identity of Indiscernibles
3.	$d(x, y) = d(y, x)$	symmetry
4.	$d(x, z) \leq d(x, y) + d(y, z)$	subadditivity or trangle inequality

1.3.2 完备（度量）空间(Complete metric space)(by mg: 柯西序列收敛于此)

[7] A metric space M is called complete (or a Cauchy space) if every **Cauchy sequence** of points in M has a limit that is also in M or, alternatively, if every Cauchy sequence in M converges in M.

Intuitively, a space is complete if there are no "points missing" from it (inside or at the boundary). For instance, the set of rational numbers is not complete, because e.g. $\sqrt{2}$ is "missing" from it, even though one can construct a Cauchy sequence of rational numbers that converges to it. It is always possible to "fill all the holes", leading to the completion of a given space.

1.4 范函

1.4.1 范数 (Norm)(by mg:向量的量化)

[8] A norm is a function that assigns a strictly positive length or size to each vector in a vector space—save for the zero vector, which is assigned a length of zero. A **seminorm**, on the other hand, is allowed to assign zero length to some non-zero vectors (in addition to the zero vector).

Given a vector space V over a subfield F of the complex numbers, a norm on V is a function $p : V \to R$ with the following properties:

For all $a \in F$ and all $u, v \in V$,

1. $p(a\boldsymbol{v}) = |a|p(\boldsymbol{v})$, (absolute homogeneity or absolute scalability).
2. $p(\boldsymbol{u} + \boldsymbol{v}) \leq p(\boldsymbol{u}) + p(\boldsymbol{v})$ (triangle inequality or subadditivity).
3. If $p(\boldsymbol{v}) = 0$ then $\boldsymbol{v}$ is the zero vector (separates points). 例：

- Absolute-value norm : $\|x\| = |x|$
- Euclidean norm (L^2 norm) : $\|x\| : \sqrt{x_1^2 + \cdots + x_n^2}$
- Taxicab norm or Manhattan norm (l_1 norm): $\|x\|_1 := \sum_{i=1}^{n} |x_i|$
- p-norm : $\|x\|_p := (\sum_{i=1}^{n} |x_i|^p)^{\frac{1}{p}}$

  The partial derivative of the p-norm

$$\frac{\partial}{\partial x_k}\|x\|_p = \frac{x_k|x_k|^{p-2}}{\|x_k\|_p^{p-1}}$$

  The derivative with respect to x

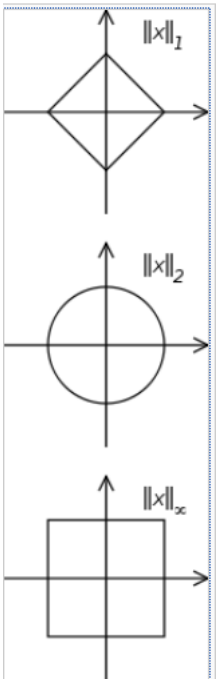
$$\frac{\partial\|x\|_p}{\partial x} = \frac{x \circ |x_k|^{p-2}}{\|x\|_p^{p-1}}$$

where $\circ$ denotes Hadamard product and $|\cdot|$ is used for absolute value of each component of the vector.

- Maximum norm (special case of: infinity norm, uniform norm, or supremum norm) If $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$,th

$$\|x\|_\infty := max(|x_1|, \ldots, |x_n|)$$

The concept of **unit circle** (the set of all vectors of norm 1) is different in different norms: for the 1-norm the unit circle in R^2 is a square, for the 2-norm (Euclidean norm) it is the well-known unit circle, while for the infinity norm it is a different square. For any p-norm it is a superellipse (with congruent axes).



1.4.2 向量空间 (vector space)

[9] A vector space (also called a **linear space**) is a collection of objects called **vectors**, which may be added together and multiplied ("scaled") by numbers, called **scalars** in this context.

A vector space over a field F is a set V together with two operations that satisfy the eight axioms listed below. Elements of V are commonly called **vectors**. Elements of F are commonly called **scalars**.

1	Associativity of addition（结合律）	$\boldsymbol{u} + (\boldsymbol{v} + \boldsymbol{w}) = (\boldsymbol{u} + \boldsymbol{v}) + \boldsymbol{w}$
2	Commutativity of addition（交换律）	$\boldsymbol{u} + \boldsymbol{v} = \boldsymbol{v} + \boldsymbol{u}$
3	Identity element of addition	There exists an element $0 \in V$, called the **zero vector**, such that $\boldsymbol{v} + 0 = \boldsymbol{v}$ for all $v \in V$.
4	Inverse elements of addition	For every $\boldsymbol{v} \in V$, there exists an element -$\boldsymbol{v} \in V$, called the **additive inverse** of $\boldsymbol{v}$, such that $\boldsymbol{v} + (\text{-}\boldsymbol{v}) = 0$.
5	Compatibility of scalar multiplication with field multiplication	$a(b\boldsymbol{v}) = (ab)\boldsymbol{v}$
6	Identity element of scalar multiplication	$1\boldsymbol{v} = \boldsymbol{v}$, where 1 denotes the multiplicative identity in F.
7	Distributivity of scalar multiplication with respect to vector addition	$a(\boldsymbol{u} + \boldsymbol{v}) = a\boldsymbol{u} + a\boldsymbol{v}$
8	Distributivity of scalar multiplication with respect to field addition	$(a + b)\boldsymbol{v} = a\boldsymbol{v} + b\boldsymbol{v}$

向量空间的例子：

- **coordinate space**

  usually denoted F^n.A vector space composed of n-tuples of a field F,$(a_1, a_2, \ldots, a_n)$ where a_i is an element of F

- **Complex numbers and other field extensions**

- **Function spaces**

  定义了函数加法:$(f+g)(w) = f(w) + g(w)$

- **Linear equations**

1.4.3 点积（dot product)(by mg:2个序列的量化)

[?] the dot product or scalar product is an algebraic operation that takes two equal-length sequences of numbers (usually coordinate vectors) and returns **a single number**.

Algebraic definition:

$$A \cdot B = \sum_{i=1}^{n} A_i B_i = A_1 B_1 + A_2 B_2 + \cdots + A_n B_n$$

例如：

$$\begin{aligned}[1,2,-5] \cdot [4,-2,-1] &= (1)(4) + (3)(-2) + (-5)(-1) \\ &= 4 - 6 + 5 \\ &= 3\end{aligned}$$

Geometric definition:

$$A \cdot B = \|A\|\|B\| cos(\theta)$$

由此有以下重要结论
若A和B正交(orthogonal)，则$A \cdot B = 0$
另外，$A \cdot A = \|A\|^2$
则，$\|A\| = \sqrt{A \cdot A}$
另外，向量A和向量B的scalar projection(or scalar component)

$$A_B = \|A\| cos\theta$$

In terms of the geometric definition of the dot product, this can be rewritten

$$A_B = A \cdot \widehat{B}$$

where $\widehat{B} = B/\|B\|$,is the unit vector in the direction of B.

因此，点积可表示为：$A \cdot B = A_B \|B\| = B_A \|A\|$

1.4.4 内积(inner product)

[?] The field of scalars denoted F is either the field of real numbers R or the field of complex numbers C.

Formally, an inner product space is a vector space V over the field F together with an inner product, i.e., with a map

$$\langle \cdot, \cdot \rangle : V \times V \to F$$

that satisfies the following three axioms for all vectors $x, y, z \in V$ and all scalars $a \in F$:

- Conjugate symmetry

$$\langle x, y\rangle = \overline{\langle y, x\rangle}$$

- Linearity in the first argument

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

- Positive-definiteness

$$\langle x, x\rangle \geq 0$$
$$\langle x, x\rangle = 0 \Leftrightarrow x = 0$$

范数相关：

- 定义内积空间的范数: $\|x\| = \sqrt{\langle x, x\rangle}$
- Cauchy–Schwarz inequality: $|\langle x, y\rangle| \leq \|x\| \cdot \|y\|$
- angel: $angle(x, y) = arccos\frac{\langle x,y\rangle}{\|x\|\cdot\|y\|}$
- Orthogonal正交: 内积为0 , $\langle x, y\rangle = 0$
- Homogeneity同质性: for x an element of V and r a scalar , $\|r \cdot x\| = |r| \cdot \|x\|$
- Triangle inequality: $\|x + y\| \leq \|x\| + \|y\|$
- Pythagorean theorem: Whenever x, y are in V and $\underline{\langle x, y\rangle = 0}$, then, $\|x\|^2 + \|y\|^2 = \|x + y\|^2$

1.4.5 内积空间 (inner product space)(by mg:定义了内积的向量空间)

[?] An inner product space is a vector space with an additional structure called an **inner product**. This additional structure associates each pair of vectors in the space with a scalar quantity known as the inner product of the vectors.由此，可以定义向量的长度和向量间的夹角，以及向量正交等概念。

A complete space with an inner product is called a **Hilbert space**.

1.4.6 欧几里德空间 (Euclidean space)

[?] Euclidean space consisting of three-dimensional vectors, denoted by R^3, and equipped with the dot product. The dot product takes two vectors x and y, and produces a real number $x \cdot y$. If x and y are represented in Cartesian coordinates, then the dot product is defined by

$$(x_1, x_2, x_3) \cdot (y_1, y_2, y_3) = x_1y_1 + x_2y_2 + x_3y_3$$

1.4.7 希尔伯特空间 (Hilbert space)

[?] The dot product satisfies the properties:

1.It is **symmetric** in x and y:

$$x \cdot y = y \cdot x$$

2.It is **linear** in its first argument:

$$(ax_1 + bx_2) \cdot y = ax_1 \cdot y + bx_2 \cdot y$$

for any scalars a, b, and vectors x_1, x_2, and y.

3.It is **positive definite**: for all vectors x:

$$x \cdot x \geq 0$$

with equality if and only if $x = 0$.

by mg:在这里，正定，在内积空间被定义为向量的内积 ≥ 0

Every finite-dimensional inner product space is also a Hilbert space.

向量长度length定义为范数$\|x\|$，和角度的关系：

$$x \cdot y = \|x\|\|y\|cos\theta$$

1.5 核

1.5.1 Kernel trick

[?] Kernel methods can be thought of as instance-based learners. rather than learning some fixed set of parameters corresponding to the features of their inputs, they instead "remember" the i-th training example (x_i, y_i) and learn for it a corresponding weight w_i. Prediction for unlabeled inputs, i.e., those not in the training set, is treated by the application of a similarity function k, called a **kernel**, between the unlabeled input x' and each of the training inputs x_i. For instance, a kernelized binary classifier typically computes a weighted sum of similarities

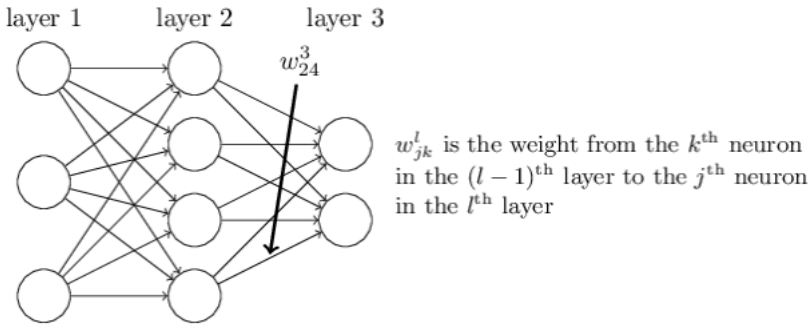
Chapter 2

Backpropagation

[?]

2.1 定义

b_j^l : l^{th} 层 j^{th} neuron 的偏置bias.
a_j^l : l^{th} 层 j^{th} neuron 的激活activation.
w_{jk}^l : l^{th}层的 node 的权重, l^{th} 层的 j^{th} neuron $\Leftarrow$ $(l-1)^{\text{th}}$ 层的k^{th} neuron 如图 :



$$a_j^l = \sigma\left(\sum_k w_{jk}^l a_k^{l-1} + b_j^l\right)$$

向量表示:

$$a^l = \sigma(w^l a^{l-1} + b^l)$$

矩阵表示 :

$$a^l = \sigma(z^l)$$
$$z^l \equiv w^l a^{l-1} + b^l$$

其中 :

$$z_j^l = \sum_k w_{jk}^l a_k^{l-1} + b_j^l$$

2.2 cost function的两个假设

后向传播的目的是得到代价函数C对于网络中任意权重w或偏差b的偏导数 :

$$\frac{\partial C}{\partial w} \text{ 和 } \frac{\partial C}{\partial b}$$

二次代价函数的形式 :

$$C = \frac{1}{2n}\sum_x \|y(x) - a^L(x)\|^2$$

n为样本数 ; $y = y(x)$为x对应的期望输出 ; L为网络层数 ; $a^L = a^L(x)$是输入为x时的网络激活输出向量。

关于代价函数，什么样的假设可以使得后向传播能够应用起来?

第一个假设是代价函数能够被写成基于每一个独立的训练样本x求代价函数C_x的平均值: $C_x = \frac{1}{n}\sum_x C_x$。二次代价函数满足此条件，其中每一个样本的代价为 : $C_x = \frac{1}{2}\|y - a^L\|^2$

需要这个假设的原因是因为后向传播实际上让我们能够基于每一个样本计算偏导数 $\partial C_x/\partial w$ 和 $\partial C_x/\partial b$ 我们然后在整个选练样本基础上经过平均而求出 $\partial C/\partial w$ 和 $\partial C/\partial b$。

第二个关于代价函数的假设是可以把它当作神经网路激活输出的一个函数， $cost\ C = C(a^L)$ 二次型代价函数满足这个假定，因为对于每一个样本x，代价函数可以写成:

$$C = \frac{1}{2}\|y - a^L\|^2 = \frac{1}{2}\sum_j (y_j - a_j^L)^2$$

Hadamard乘积或Schur乘积，$\odot$，即逐元素相乘

2.3 backpropagation背后的基础四等式

引入中间变量δ_j^l，为做网络中第l^{th}层第j^{th}神经元的误差: $\delta_j^l \equiv \frac{\partial C}{\partial z_j^l}$

2.3.1 输出层的误差等式:

$$\delta_j^L = \frac{\partial C}{\partial a_j^L}\sigma'(z_j^L). \tag{BP1}$$

矩阵形式：

$$\delta^L = \nabla_a C \odot \sigma'(z^L). \tag{BP1a}$$

2.3.2 误差δ^l用下一层的误差δ^{l+1}表示:

$$\delta^l = ((w^{l+1})^T\delta^{l+1}) \odot \sigma'(z^l), \tag{BP2}$$

2.3.3 代价函数相对网络中任意偏置变化率的等式:

$$\frac{\partial C}{\partial b_j^l} = \delta_j^l. \tag{BP3}$$

误差δ_j^l和变化率$\partial C/\partial b_j^l$精确相同，简短形式为：

$$\frac{\partial C}{\partial b} = \delta, \tag{31}$$

2.3.4 代价函数相对网络中任意权重变化率的等式:

$$\frac{\partial C}{\partial w_{jk}^l} = a_k^{l-1}\delta_j^l. \tag{BP4}$$

改写为：

$$\frac{\partial C}{\partial w} = a_{\text{in}}\delta_{\text{out}}, \tag{32}$$

可以看到，(BP4) 的结果就是小的激活神经元的权重会比较缓慢的学习。

summary: backpropagation 方程

$$\delta^L = \nabla_a C \odot \sigma'(z^L). \tag{BP1}$$
$$\delta^l = ((w^{l+1})^T \delta^{l+1}) \odot \sigma'(z^l), \tag{BP2}$$
$$\frac{\partial C}{\partial b^l_j} = \delta^l_j. \tag{BP3}$$
$$\frac{\partial C}{\partial w^l_{jk}} = a^{l-1}_k \delta^l_j. \tag{BP4}$$

矩阵形式：

（BP1)重写为

$$\delta^L = \Sigma'(z^L)\nabla_a C, \tag{1}$$

其中，$\Sigma'(z^L)$是一个对角线元素为$\sigma'(z^L_j)$的方阵，且其非对角线元素都是零。 此矩阵与$\nabla_a C$进行传统的矩阵乘法

(BP2)重写为：

$$\delta^l = \Sigma'(z^l)(w^{l+1})^T \delta^{l+1} \tag{2}$$

合并（1）和（2）可得：

$$\delta^l = \Sigma'(z^l)(w^{l+1})^T \dots \Sigma'(z^{L-1})(w^L)^T \Sigma'(z^L)\nabla_a C \tag{3}$$

2.4 The Backpropagation

1. Input x:
Set the corresponding activation a^1 for the input layer.

2. Feedforward:
For each $l = 2, 3, \dots, L$ compute
$$z^l = w^l a^{l-1} + b^l \text{ and } a^l = \sigma(z^l).$$

3. Output error δ^L:
Compute the vector
$$\delta^L = \nabla_a C \odot \sigma'(z^L).$$

4. Backpropagate the error:
For each $l = L-1, L-2, \dots, 2$ compute
$$\delta^l = ((w^{l+1})^T \delta^{l+1}) \odot \sigma'(z^l).$$

5. Output: The gradient of the cost function
$$\frac{\partial C}{\partial w^l_{jk}} = a^{l-1}_k \delta^l_j \text{ and } \frac{\partial C}{\partial b^l_j} = \delta^l_j.$$

单个样本训练过程：

1. 输入训练样本集合
2. 对于每一个训练样本x：

设置对应的输入激活 $a^{x,1}$执行以下步骤：

○向前：对于每一层$l = 2, 3, \ldots, L$,计算

$$z^{x,l} = w^l a^{x,l-1} + b^l \text{ 和 } a^{x,l} = \sigma(z^{x,l})$$

○输出层误差$\delta^{x,L}$：计算向量

$$\delta^{x,L} = \nabla_a C_x \odot \sigma'(z^{x,L}).$$

○后向传播误差：对于每一层$l = L-1, L-2, \ldots, 2$，计算

$$\delta^{x,l} = ((w^{l+1})^T \delta^{x,l+1}) \odot \sigma'(z^{x,l}).$$

3. 梯度下降：按下面规则更新，

$$w_k \longrightarrow w'_k = w_k - \eta \frac{\partial C_x}{\partial w_k} \tag{2.1}$$

$$b_l \longrightarrow b'_l = b_l - \eta \frac{\partial C_x}{\partial b_l} \tag{2.2}$$

即，对于每一层$l = L, L-1, \ldots, 2$,更新为：

$$w^l \to w^l - \frac{\eta}{m} \sum_x \delta^{x,l} (a^{x,l-1})^T, \tag{2.3}$$

$$b^l \to b^l - \frac{\eta}{m} \sum_x \delta^{x,l} \tag{2.4}$$

Chapter 3

Transfer Learning

3.1 RKHS

[?] In functional analysis (a branch of mathematics), a **reproducing kernel Hilbert space (RKHS)** is a Hilbert space of functions in which point evaluation is a continuous linear functional.

Roughly speaking, this means that if two functions f and g in the RKHS are close in norm, i.e., $\|f-g\|$ is small, then f and g are also pointwise close, i.e., $|f(x)-g(x)|$ is small for all x. The reverse need not be true.

Let X be an arbitrary set and H a Hilbert space of real-valued functions on X. The evaluation functional over the Hilbert space of functions H is a linear functional that evaluates each function at a point x,

$$L_x : f \mapsto f(x) \forall f \in H$$

We say that H is a reproducing kernel Hilbert space if L_x is continuous at any f in H or, equivalently, if for all x in X, L_x is a bounded operator on H , i.e. there exists some $M > 0$ such that

$$L_x[f] := f(x) \leq M\|f\|_H \forall f \in H$$

3.2 MMD

[?] define the maximum mean discrepancy (MMD) as:

p and q be distributions defined on a domain X

obervations: $X := \{x_1, \ldots, x_m\}$, $Y := \{y_1, \ldots, y_n\}$,drawn independently and identically distributed (i.i.d.) from p and q respectively

$$MMD[\mathcal{F}, p, q] := \sup_{f \in \mathcal{F}} \left(E_{x\sim p}[f(x)] - E_{y\sim p}[f(y)]\right) \tag{3.1}$$

$$MMD[\mathcal{F}, X, Y] := \sup_{f \in \mathcal{F}} \left(\frac{1}{m}\sum_{i=1}^{m} f(x_i) - \frac{1}{n}\sum_{i=1}^{n} f(y_i)\right) \tag{3.2}$$

Since $E_p[f(x) = \langle \mu[p], f\rangle]$,we may rewrite

$$MMD[\mathcal{F}, p, q] = \sup_{\|f\|_{\mathcal{H}} \leq 1} \langle \mu[p] - \mu[q], f\rangle = \|\mu[p] - \mu[q]\|_{\mathcal{H}}$$

using $\mu[X] := \frac{1}{m}\sum_{i=1}^{m}$ and $k(x, x') = \langle \phi(x), \phi(x')\rangle$ empirical estimate of MMD is :

$$MMD[\mathcal{F}, X, Y] = \left[\frac{1}{m^2}\sum_{i,j=1}^{m} k(x_i, y_j) - \frac{2}{mn}\sum_{i,j=1}^{m,n} k(x_i, y_j) + \frac{1}{n^2}\sum_{i,j=1}^{n} k(y_i, y_j)\right]^{\frac{1}{2}}$$

3.3 Deep Transfer Network

[?]

Denote $X^s = [x_1^s, \ldots, x_{n^s}^s] \in R^{d \times n^s}$ and $X^t = [x_1^t, \ldots, x_{n^t}^t] \in R^{d \times n^t}$ as the data matrices of D^s and D^t respectively,$X = [X^s, X^t]$ as the combination of X^s and X^t

$$MMD = \| \frac{1}{n^s} \sum_{i=1}^{n^s} x_i^s - \frac{1}{n^t} \sum_{j=1}^{n^t} x_j^t \|_2^2 \tag{3.3}$$

$$= Tr(XMX^T) \tag{3.4}$$

where M is the MMD matrix. Let M_{ij} be one element of M.

$$M_{ij} = \{ \tag{3.5}$$

Chapter 4

whiten

Bibliography

[1] "Infimum and supremum - Wikipedia, the free encyclopedia." bibtex: _infimum_????

[2] "Cauchy sequence - Wikipedia, the free encyclopedia." bibtex: _cauchy_????

[3] "Positive-definite kernel - Wikipedia, the free encyclopedia." bibtex: _positive-definite_????-2.

[4] "Topology - Wikipedia, the free encyclopedia." bibtex: _topology_????

[5] "Topological vector space - Wikipedia, the free encyclopedia." bibtex: _topological_????

[6] "Metric space - Wikipedia, the free encyclopedia." bibtex: _metric_????

[7] "Complete metric space - Wikipedia, the free encyclopedia." bibtex: _complete_????

[8] "Norm (mathematics) - Wikipedia, the free encyclopedia." bibtex: _norm_????

[9] "Vector space - Wikipedia, the free encyclopedia." bibtex: _vector_????