Тема 1

Сведения из высшей математики

В дальнейшем нам понадобятся некоторые сведения из алгебры и методов оптимизации. Поэтому далее приводятся основные понятия и сведения из этих дисциплин. Читатели знакомые с этими сведениями могут пропустить раздел и перейти к следующей теме.

Традиционно используемым для описания нейронных сетей математическим языком является аппарат векторной и матричной алгебры.

Векторные пространства

Основным структурным элементом в описании способов обработки информации нейронной сетью является вектор - упорядоченный набор чисел, называемых компонентами вектора. В дальнейшем вектора будут обозначаться латинскими буквами $(\bar{a},\bar{b},\bar{c},\bar{d})$, а скаляры - числа - греческими буквами $(\alpha,\beta,\gamma,\delta)$. Для обозначения матриц будут применяться заглавные латинские буквы. Компоненты вектора могут быть действительными числами, целыми числами, булевыми числами. Компоненты вектора $\bar{x}=(x_1,x_2,x_3,\ldots,x_n)$ можно рассматривать, как его координаты в некотором n-мерном пространстве. В случае действительных компонент это пространство обозначается, как R^n и включает в себя набор всех возможных совокупностей из n действительных чисел. В случае двоичных компонент это пространство будем обозначать B^n. Говорят, что вектор $\bar{x}$ принадлежит пространству R^n $(\bar{x}\in R^n)$. В дальнейшем, если нам потребуется набор векторов, мы будем нумеровать их верхними индексами, чтобы не путать с нумерацией компонент вектора: $\{\bar{x}^1,\bar{x}^2,\ldots,\bar{x}^k\}$.

Определение 1.1. Линейное векторное пространство - множество векторов V, для которых определены операции векторного сложения и умножения на вещественное число, удовлетворяющие перечисленным ниже соотношениям.

Для любых $x,y,z\in V$; $a,b\in R$

1) переместительный закон сложения: $x+y=y+x$;
2) сочетательный закон сложения: $(x+y)+z=x+(y+z)$;
3) существование нулевого элемента: Если $0\in V \quad x\in V \Rightarrow 0+x=x\in V$;
4) существование противоположного элемента: $\forall x\in V$ существует $-x\in V$, такой, что $x+(-x)=0$

5) $1 \cdot x = x$,

6) Сочетательность умножения на скаляр: $(a \cdot b) \cdot x = a \cdot (b \cdot x)$;

7) Распределительный закон: $a \cdot (x + y) = a \cdot x + a \cdot y, (a + b) \cdot x = a \cdot x + b \cdot x$.

Элементами линейного пространства являются векторы.

Примером линейного векторного пространства является пространство R^n с покомпонентными операциями сложения и умножения.

Определение 1.2. Для двух элементов векторного пространства может быть определено их *скалярное (внутреннее) произведение*. Скалярное произведение - функция обладающая следующими свойствами:

1. $(x, y) = (y, x)$ - симметричность,
2. $(a \cdot x, y) = a \cdot (x, y)$ - линейность,
3. $(x + y, z) = (x, z) + (y, z)$ - аддитивность по каждому сомножителю, а также
4. $(x, x) \geq 0$, причем $(x, x) = 0 \Leftrightarrow x = 0$ - неотрицательность.

Линейное векторное пространство, в котором определено скалярное произведение, удовлетворяющее аксиомам 1-4, называется *евклидовым,* его мы будем обозначать буквой *E*.

Пример 1.1. Определим скалярное произведение двух действительных векторов *x*, *y* так:

$$f(x,y) = (x, y) = x_1 \cdot y_1 + x_2 \cdot y_2 + \ldots + x_n \cdot y_n$$

Равенство нулю скалярного произведения двух векторов означает их взаимную ортогональность, сообразно обычным геометрическим представлениям.

Для векторов можно ввести понятие *нормы* – длины вектора. Пространство, в котором определена норма векторов, называется *нормированным*.

Определение 1.3. Норма в пространстве *E* - это функция, определенная для каждого вектора пространства, обозначаемая символом $|\ |$ и обладающая следующими свойствами:

$\forall x, y \in E$

1. $|x| \geq 0$, причем $|x| = 0 \Leftrightarrow x = 0$;
2. $|a \cdot x| = |a| \cdot |x|$;
3. $|x + y| \leq |x| + |y|$.

Пример 1.2. Для евклидова пространства можно ввести норму как $|x| = \sqrt{(x, x)}$.

Для векторов, состоящих из действительных чисел, мы будем в дальнейшем иметь дело именно с евклидовым пространством. В случае булевых векторов размерности n рассматриваемое пространство представляет собой множество вершин n-мерного гиперкуба с Хемминговой метрикой. Расстояние между двумя вершинами определяется длиной кратчайшего соединяющего их пути, измеренной вдоль ребер:

$$d_H(x,y)=\sum_{i=1}^{n}(x_i \oplus y_i)\,,\ \ x,y \in V^n=\{0,1\}^n\,.$$

Евклидова метрика для прямоугольной системы координат определяется формулой:

$$d_E(x,y)=|x-y|=\sqrt{(x-y,x-y)}=\sqrt{(x_1-y_1)^2+(x_2-y_2)^2+\ldots+(x_n-y_n)^2}\,,\ \ x,y\in R^n$$

Расстояние в данном случае может трактоваться как мера сходства двух образов.

Еще одно важное понятие, которое понадобиться в дальнейшем – это угол между двумя векторами. Углом между векторами *x* и *y* назовём число

$$\varphi=\arccos\frac{(x,y)}{|x|\cdot|y|}\,.$$

Вектора, для которых $\varphi=\pi$, то есть *(x,y)=0* будем называть *ортогональными*.

Базис линейного векторного пространства

Определение 1.4. Вектора $x^1,x^2,\ldots,x^m$ считаются *линейно независимыми*, если их произвольная линейная комбинация $a_1x^1+a_2x^2+\ldots+a_mx^m$ не обращается в ноль, если только все константы $a_1,\ldots,a_m$ не равны одновременно нулю.

Определение 1.5. В линейном векторном пространстве L размерности n можно подобрать такой минимальный набор векторов $\{e^1,e^2,\ldots,e^n\}$, что любой вектор $x\in L$ можно представить в виде линейной комбинации векторов этого набора. Такой набор векторов называется *базисом*. Можно доказать, что базис может состоять из любой комбинации из n линейно независимых векторов, где n - размерность пространства.

Выберем некоторую систему линейно независимых векторов $x^1,x^2,\ldots,x^m,\quad m<n$. Все возможные линейные комбинации этих векторов сформируют линейное пространство размерности m, которое будет являться *подпространством* или *линейной оболочкой* L исходного n-мерного пространства. Выбранная базовая система из m векторов является, очевидно, базисом в подпространстве L. Важным частным случаем линейной оболочки является подпространство размерности на единицу меньшей, чем размерность исходного пространства, называемое *гиперплоскостью*. В случае трехмерного пространства это обычная плоскость. Гиперплоскость делит пространство на две части. Совокупность

гиперплоскостей разбивает пространство на несколько множеств, каждое из которых содержит вектора с близким набором признаков, тем самым может осуществляться классификация векторов.

Для двух подпространств может быть введено понятие их взаимной ортогональности. Два подпространства L_1 и L_2 называются взаимно ортогональными, если всякий элемент одного подпространства ортогонален каждому элементу второго подпространства.

Ортогонализация набора векторов

Произвольно выбранные линейно независимые вектора необязательно являются взаимно ортогональными. Однако в ряде приложений удобно работать с ортогональными системами. Для этого исходные вектора требуется ортогонализовать. Классический процесс ортогонализации Грама-Шмидта состоит в следующем: по системе линейно независимых ненулевых векторов $x^1, x^2, \ldots, x^m$ рекуррентно строится система ортогональных векторов $h^1, h^2, \ldots, h^m$. В качестве первого вектора h^1 выбирается исходный вектор x^1. Каждый следующий (i-ый) вектор делается ортогональным всем предыдущим, для чего из него вычитаются его проекции на все предыдущие вектора:

$$h^i = x^i - \sum_{j=1}^{i-1} \frac{(x^i, h^j)}{\left|h^j\right|^2} h^j .$$

При этом, если какой-либо из получившихся векторов h^i оказывается равным нулю, он отбрасывается. Можно показать, что, по построению, полученная система векторов оказывается ортогональной, т.е. каждый вектор содержит только уникальные для него признаки.

Матрицы и линейные преобразования векторов

Равно тому, как был рассмотрен вектор - объект, определяемый одним индексом (номером компоненты или признака), может быть введен и объект с двумя индексами, *матрица*. Эти два индекса определяют компоненты матрицы A_{ij}, располагаемые по строкам и столбцам, причем первый индекс i определяет номер строки, а второй j - номер столбца. В некоторых случаях упорядоченный набор элементов может трактоваться и как вектор и как матрица (например, матрица размерности 5x5 может рассматриваться как вектор, состоящий из двадцати пяти компонент).

Суммой двух матриц A и B одинаковой размерности $(n \times m)$ является матрица C той же размерности с компонентами, равными сумме соответствующих компонент исходных матриц: $C_{ij} = A_{ij} + B_{ij}$. Матрицу можно умножить на скаляр, при этом в результате получается матрица той же размерности, каждая компонента которой умножена на этот скаляр. Произведением двух матриц $A(n \times l)$ и $B(l \times m)$ является матрица $C(n \times m)$, компоненты которой даются соотношением:

$$C_{ij} = \sum_{k=1}^{l} A_{ik} B_{kj} .$$

Заметим, что размерности перемножаемых матриц должны быть согласованными - число столбцов первой матрицы должно равняться числу строк второй.

В важном частном случае, когда вторая матрица является вектором (т.е. матрицей с одной из размерностей, равной единице (*m=1*)), представленное правило определяет способ умножения матрицы на вектор:

$$c_i = \sum_{k=1}^{l} A_{ik} b_k .$$

В результате умножения получается также вектор $\bar{c}$, причем для квадратной матрицы $A(l \times l)$, его размерность равна размерности вектора-сомножителя $\bar{b}$. При произвольном выборе квадратной матрицы A можно построить произвольное линейное преобразование одного вектора $(\bar{x})$ в другой $(\bar{y})$ той же размерности: $\bar{y} = A\bar{x}$. Более точно, для того, чтобы преобразование T одного вектора в другой являлось линейным, необходимо и достаточно, чтобы для двух векторов $\bar{x}^1$ и $\bar{x}^2$ и чисел a и b выполнялось равенство: $T(a\bar{x}^1 + b\bar{x}^2) = aT(\bar{x}^1) + bT(\bar{x}^2)$. Можно показать, что всякому линейному преобразованию векторов соответствует умножение исходного вектора на некоторую матрицу.

Если в приведенной выше формуле для умножения матрицы A на вектор $\bar{x}$ компоненты этого вектора неизвестны, в то время, как A и результирующий вектор b известны, то о выражении $A\bar{x} = b$ говорят, как о системе линейных алгебраических уравнений относительно компонент вектора $\bar{x}$. Система имеет единственное решение, если векторы, определяемые строками квадратной матрицы A, являются линейно независимыми. В этом случае также говорят что матрица *A невырождена.*

Часто используемыми частными случаями матриц являются диагональные матрицы, у которых все элементы вне главной диагонали равны нулю. Диагональную матрицу, все элементы главной диагонали которой равны единице, называют единичной матрицей I. Линейное преобразование, определяемое единичной матрицей, является тождественным: $I\overline{x} = \overline{x}$ для всякого вектора $\overline{x}$.

Для матриц определена, кроме операций умножения и сложения, также операция транспонирования. Транспонированная матрица A^T получается из исходной матрицы A заменой строк на столбцы: $A_{ij}^T = A_{ji}$. Матрицы, которые не изменяются при транспонировании, называют симметричными матрицами. Для компонент симметричной матрицы S имеет место соотношение $S_{ij} = S_{ji}$. Всякая диагональная матрица, очевидно, является симметричной.

Пространство квадратных матриц одинаковой размерности с введенными операциями сложения и поэлементного умножения на скаляр, является линейным пространством. Для него также можно ввести метрику и норму. Нулевым элементом служит матрица, все элементы которой равны нулю.

В заключении приведем некоторые тождества для операций над матрицами. Для всяких A, B и C и единичной матрицы I имеет место:

$$IA = AI = A,$$
$$(AB)C = A(BC),$$
$$A(B + C) = AB + AC,$$
$$(A^T)^T = A,$$
$$(A + B)^T = A^T + B^T,$$
$$(AB)^T = B^T A^T.$$

Литература

1. Курош А. Г. Лекции по общей алгебре. «Наука», Москва, 1973.
2. В. А. Ильин, Э. Г. Позняк. Линейная Алгебра, Москва, Наука, 1999.