Машинное обучение в финансах

Лекция 2. Математика: Функции, производная, интеграл. Линейная алгебра.

Роман В. Литвинов*

*CRO
Финансовая Группа БКС

Высшая школа экономики, Январь 2021

Содержание

Содержание

Язык

Математика в первую очередь будет интересовать нас как универсальный язык для эффективной коммуникации:

- коммуницирования своих идей
- эффективного восприятия чужих идей

NB: Мы не собираемся фокусироваться на доказательстве теорем, лемм и проч. Это предмет другого курса (research vs implementation)

По ходу курса мы также увидим как одни и те же объекты (множества, функции, матрицы и тп), которые мы видим в математике, "физически"реализуются в программировании и машинном обучении.

NB: Подразумевается, что все что мы будем обсуждать, вы уже знаете к этому моменту. Но повторить будет не лишним. Плюс, см. тезис выше.

Язык. Множества

Множество (set)

Множество – совокупность объектов, которые называются элементами этого множества.

Примеры: множество акций, торгуемых на MICEX. Множество характеристик финансового инструмента. Множество натуральных чисел (N).

Пустое множество (empty set)

Множество не имеещее элементов. Обозначается $\varnothing$.

Пусть X - некоторе множество и a - произвольный элемент. Тот факт, что a принадлежит множеству X (входит в состав элементов X) обозначается как a$\in \mathrm{X}$.

В случае если a не принадлежит X (не явл. элементом X) пишут a$\notin \mathrm{X}$

Язык. Множества

Множества равны, если они состоят из одинаковых элементов.

Подмножество (subset)

Множество A называется подмножеством множества B, если все его элементы содержатся в B. Обозначается A$\subseteq \mathrm{B}$.

Пусть A и B два множества. Их объединение - это множество состоящее из всех элементов принадлежащих A или B. Обозначается A$\cup \mathrm{B}$.

Пересечение - это множество всех элементов принадлежащих A и B. Обозначается A$\cap \mathrm{B}$.

Примеры: "склейка"двух массивов данных (дублирующие элементы отбрасываются). Множество дублирующихся значений в двух (и более) массивах данных.

Язык. Множества. Функции

Функция (function)

Пусть X и Y два множества. Функцией из X в Y называется правило, которое ставит в соответствие каждому элементу $x \in X$ элемент $y \in Y$ и при том единственный. Обозначается $f : X \to Y$.

При заданном элементе $x \in X$ выражение $f(x)$ обозначает элемент $y \in Y$, который соотвествует данному x (значение функции на элементе x).

$$X \to \boxed{f} \to Y$$

Множество X называется областью определения функции, x - аргументом функции или независимой переменной. Множество Y - множество значений, y - зависимая переменная или значение функции.

Пример: Формула длины окружности $L = 2\pi r$

Пример: Формула объема шара $V = 4/3\pi r^3$

Пример: Формула Блэка-Шоулза для оценки стоимости Европейского опциона (функция нескольких переменных)

$$f_c = S_0\Phi(d_1) - K\mathrm{e}^{-rT}\Phi(d_2),$$
$$f_p = K\mathrm{e}^{-rT}\Phi(-d_2) - S_0\Phi(-d_1),$$

где

$$d_1 = \frac{\log(S_0/K) + (r + \sigma^2/2)T}{\sigma\sqrt{T}}, \tag{1}$$

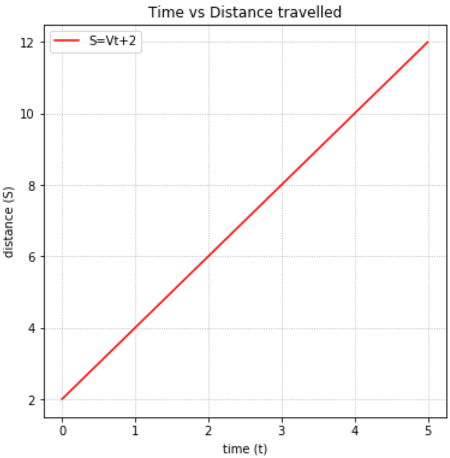
$$d_2 = \frac{\log(S_0/K) + (r - \sigma^2/2)T}{\sigma\sqrt{T}} = d_1 - \sigma\sqrt{T}. \tag{2}$$

Язык. Множества. Функции

Пример: Давайте подумаем от каких переменных/факторов могла бы зависеть стоимость акций/доходность Tesla? Apple? Тиньков? Как могла бы выглядеть подобная функция? И, кстати, как выглядит доходность, как функция от цены акции?

Содержание

Производная и ее смысл



Попробуем вычислить "мгновенную"скорость?

Производная и ее смысл

$V = (S(t_1) - S(t_3))/(t_1 - t_3)$

$V = (S(t_1) - S(t_2))/(t_1 - t_2)$

...

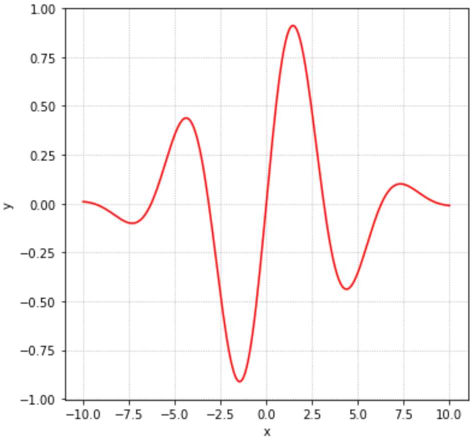
$\Delta t \to 0$ или δt

$\Delta S \to$ ... к некой величине δS

Производная (derivative)

Производная функции - предел отношения приращения функции к приращению независимой переменной при стремлении к нулю приращения независимой переменной: $\frac{\partial S}{\partial t} = \lim_{t \to 0} \frac{\Delta S}{\Delta t}$

Производная и ее смысл. Максимумы и минимумы



Точки роста/убывания функции. Точки максимума/минимума (локального/ глобального)

Производная и ее смысл. Максимумы и минимумы

Положительная производная указывает на то, что функция является растущей в данной точке, а отрицательная - на то, что убывающей. Почему?

Производная характеризует скорость роста. Она равна отношению изменения y к небольшому изменению x (чувствительность функции к "возмущениям"x).

В точках, где производная переходит от положительных значений к отрицательным функция достигает ... максимума или минимума? В точках, где производная переходит от отрицательных значений к положительным?

Приравнивая производную к нулю, можно найти значения x, в которых находятся "потенциальные"максимумы и минимумы функции (есть исключения. Например, функция $y = x^3$ с производной $y = 3x^2$. В т. $x = 0$ ее значение 0, но достаточно вспомнить график кубической параболы, чтобы понять, что это не точка максимума или минимума).

Крис Андерсон и градиентный спуск (gradient descent)

«The plume of my scent is densest next to me and disperses as it spreads, an invisible fog of particles exuded from my skin that moves like smoke with the wind. The closer to my skin, the higher the particle density; the farther away, the lower. This decrease is called a gradient, which describes any gradual transition from one level to another one—as opposed to a "step function," which describes a discrete change.

Once the mosquito follows this gradient to its source using its simple algorithm, it lands on my skin, which it senses with the heat detectors in its feet, which are attuned to another gradient—temperature. It then pushes its needle-shaped proboscis through the surface, where a third where a third set of sensors in the tip detect yet another gradient, that of blood density. This flexible needle wriggles around under my skin until the scent of blood steers it to a capillary, which it punctures. Then my blood begins to flow into the mosquito. Mission accomplished. Ouch.»

Chris Anderson "Gradient descent"("Possible minds"ed. John Brockman)

Крис Андерсон и градиентный спуск (gradient descent)

«Define a “cost function” that determines how well the network solved the problem.

Run the network once and see how it did at that cost function.

Change the values of the connections and do it again. The difference between those two results is the direction, or “slope,” in which the network moved between the two trials.

If the slope is pointed “downhill,” change the connections more in that direction. If it’s “uphill,” change them in the opposite direction. Repeat until there is no improvement in any direction. That means that you’re in a minimum.»

Chris Anderson "Gradient descent"("Possible minds"ed. John Brockman)

Содержание

Правила дифференцирования

Константа:

$$\frac{d}{dx}c = 0$$

Степень:

$$\frac{d}{dx}x^n = nx^{(n-1)}$$

Очевидно, что производная суммы (разности) двух функций равна сумме (разности) производных этих функций.

Произведение:

$$\frac{d}{dx}(f(x)g(x)) = f(x)\frac{d}{dx}g(x) + \frac{d}{dx}f(x)g(x)$$

Правила дифференцирования

Деление:

$$\frac{d}{dx}\left(\frac{f(x)}{g(x)}\right) = \frac{\frac{d}{dx}f(x)g(x) - f(x)\frac{d}{dx}g(x)}{g^2(x)}$$

Умножение на константу (следствие из правила умножения):

$$\frac{d}{dx}cf(x) = c$$

Дифференцирование сложной функции:

$$\frac{d}{dx}[f(u)] = \frac{d}{du}[f(u)]\frac{du}{dx}$$

Содержание

Частная производная

Частная производная (partial derivative) представляет собой вариант производной для функции от нескольких пременных.

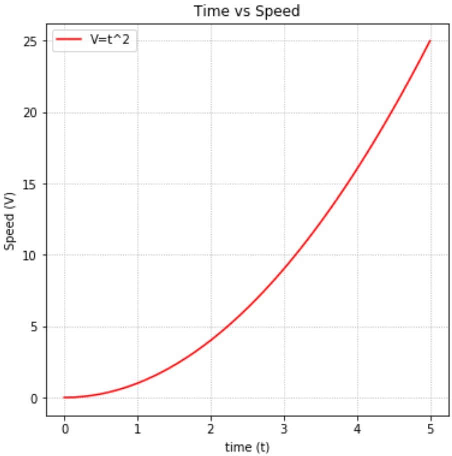
При этом мы изменяем лишь одну переменную, а остальные оставляем без изменений и получаем предел отношения приращения функции по выбранной переменной к приращению этой переменной.

$$\frac{\partial f}{\partial x_1} = \lim_{\Delta x_1 \to 0} \frac{\Delta f(x_1 + \Delta x_1, x_2, ..., x_n)}{\Delta x_1}$$

В дальнейшем нам также могут понадобиться связанные с частной производной понятия градиента (gradient), матрицы Якоби и тп. но мы будем вводить их по мере необходимости в соотвествующей части курса.

Содержание

Интеграл



Обратная задача: найти пройденный путь, если мгновенная скорость как функция от времени нам известна.

Интеграл

Рассмотрим простой случай, когда скорость V постоянна (помните наш график в разделе с производной). Тогда:

$S = (t_n - t_0)V_0$

Наш случай не такой простой и мгновенная скорость меняется в зависимости от времени. Что здесь можно придумать? Попробуем получить приблизительную оценку.

Шаг 1: разобьем наш временной отрезок на несколько непересекающихся интервалов. Например, пять: $(t_1 - t_0), ..., (t_5 - t_4)$

Шаг 2: на каждом таком отрезке будет считать скорость постоянной (это наше допущение для приблизительной оценки) и равной скорости в конце каждого отрезка: $V(t_1), ..., V(t_5)$

Шаг 3: Расчитаем путь пройденный за каждый такой отрезок: $S_1 = V(t_1)(t_1 - t_0)$

Интеграл

Шаг 4: Сложим пути пройденные за каждый из отрезков, чтобы получить общий пройденный путь. $S = S_1 + ... S_5$. Или в общем виде:

$$S(0,5) = \sum_{n=1}^{5} V_n (t_n - t_{n-1})$$

Как мы можем улучшить нашу оценку? Разбивать временной отрезок на все большее количество интервалов и использовать предложенную выше схему.

$$S(a,b) = \sum_{i=1}^{n} V_i (t_i - t_{i-1})$$

В пределе при n стремящемся к бесконечности и $\Delta t = t_i - t_{i-1}$ стремящемся к нулю, мы получим определенный интеграл.

Интеграл

Таким образом определенный интеграл в общем виде и есть:

$$\int_a^b f(x)dx = \lim_{\Delta x_1 \to 0} \sum_{i=1}^{n} f(x_i)(\Delta x_i)$$

NB: Также понятно, почему часто при упоминании оперделенного интеграла используют площади криволинейных фигур.

Функция F(x) есть неопределенный интеграл функции f(x), если

$$\frac{dF(x)}{dx} = f(x)$$

Фундаментальная теорема анализа (Теорема Ньютона-Лейбница):

$$\int_a^b f(x)dx = F(b) - F(a) \text{ где } \frac{dF(x)}{dx} = f(x)$$

Содержание

Правила интегрирования

Совпадающий верхний и нижний пределы:

$$\int\limits_a^a f(x)dx = 0$$

Перестановка местами пределов интегрирования:

$$\int\limits_b^a f(x)dx = -\int\limits_a^b f(x)dx$$

Умножение на константу:

$$\int\limits_a^b \mathrm{c}f(x)dx = \mathrm{c}\int\limits_a^b f(x)dx$$

Правила интегрирования

Сумма (аналогично для разности):

$$\int_a^b f(x)+g(x))dx = \int_a^b f(x)dx + \int_a^b g(x)dx$$

В случае, если $\mathrm{c} \in [a, b]$ и функция интегрируема на каждом из участков:

$$\int_a^b f(x)dx = \int_a^{\mathrm{c}} f(x)dx + \int_{\mathrm{c}}^b f(x)dx$$

Содержание

Вектор. Матрица

Скаляр - одно число. Например, 12.

Под вектором мы будем понимать (одномерный) массив чисел. Эти числа будут расположены в определенном порядке и каждое число будет иметь индекс соотвествующего порядка.

Пример (вектор-строка):

$$\begin{bmatrix} x_1 & x_2 & x_3 & x_4 \end{bmatrix}$$

Пример (вектор-стобец):

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$$

Вектор. Матрица

Матрица - двумерный массив чисел. Каждое число будет расположено в определенном порядке и идентифицироваться с помощью двух индексов.

$$\begin{bmatrix} x_{1,1} & x_{1,2} & x_{1,3} \\ x_{2,1} & x_{2,2} & x_{2,3} \\ x_{3,1} & x_{3,2} & x_{3,3} \end{bmatrix}$$

Матрица имеет размерность m x n в случае, если состоит из m строк и n столбцов.

В примере выше приведена матрица размерности 3 x 3.

Единичная матрица (identity matrix) I - матрица на главной диагонали которой расположены единицы, а остальные элементы равны нулю.

Диагональная матрица - матрица все элементы которой за исключением лежащих на главной диагонали, равны нулю.

Квадратная матрица - матрица количество строк которой равно количеству столбцов.

В отдельных случаях нам может понадобиться понятие тензора, и здесь (не путать с физикой и тп), мы будем просто подразумевать массив размерности более чем два.

Содержание

Матричные вычисления

Транспонирование матрицы - это замена исходных строк матрицы на столбцы.

Вектор можно представить как матрицу, имеющую один стобец (строку). В таком случае транспонирование - это преобразование вектора-строки в вектор-столбец (и наоборот).

Транспонирование матрицы A обозначается как A^T.

В случае если матрицы имеют одинаковую размерность (количество строк и столбцов совпадает) мы можем проводить сложение матриц просто складывая их соотвествующие элементы: $C = A + B$ где $C_{i,j} = A_{i,j} + B_{i,j}$.

Мы также можем умножать матрицу или вектор на скаляр. Каждый элемент в таком случае умножается на соотвествующее число.

Матричные вычисления

Произведением матрицы A $n \times p$ на матрицу B $p \times q$ называется матрица C $n \times q$ такая, что элемент матрицы C, стоящий в i-ой строке и j-ом столбце, т.е. элемент c-ij, равен сумме произведений элементов i-ой строки матрицы A на соответствующие элементы j-ого столбца матрицы B.

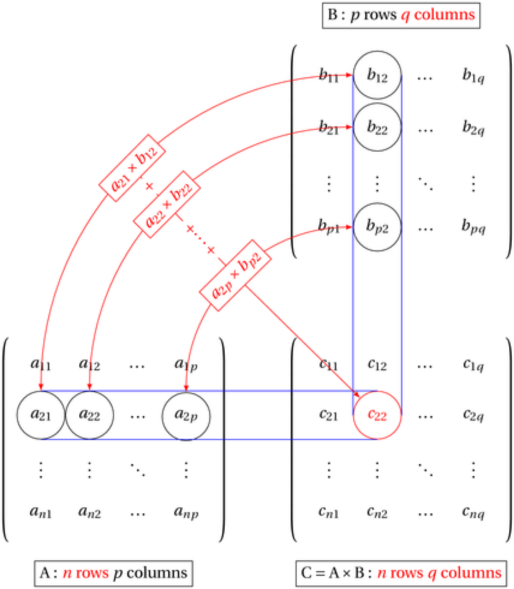
Шаг 1: убедитесь, что количество столбцов первой матрицы равно количеству строк второй матрицы (иначе умножать невозможно!)

Шаг 2: умножьте элементы i-строки первой матрицы на элементы j-ого столбца второй матрицы.

Шаг 3: сложите полученный результат и поставьте на соотвествующее место (ij).

См. ниже (Alain Matthes (c) http://altermundus.com/pages/examples.html)

Матричные вычисления



Матричные вычисления

Матричное умножение дистрибутивно: $\mathrm{A}(\mathrm{B}+\mathrm{C}) = \mathrm{AB}+\mathrm{AC}$.

Матричное умножение ассоциативно: $\mathrm{A}(\mathrm{BC}) = (\mathrm{AB})\mathrm{C}$.

В общем случае матричное умножение некоммутативно. Условие когда $\mathrm{AB} = \mathrm{BA}$ не всегда соблюдается.

Также есть смысл отметить, что операция транспонирования обладает следующим свойством: $(AB)^T = B^T A^T$

Скалярное произведение двух векторов (одинаковой размерности) в Евклидовом n-мерном пространстве (dot product) можно представить как операцию $x^T y$. Такое произведение коммутативно.

Обратной матрицей A^{-1} (inverse matrix) матрицы A, является матрица соотвествующая условию $A^{-1}A = I$.

Литература

Дополнительная литература к сегодняшней лекции:

- Brockman,J. ed. (2019) Possible minds. 25 ways looking at AI. Penguin press.
- Deisenroth, M. Fisal, A., Ong, C.S. (2020) Mathematics for machine learning. Cambridge. (Ch.2-5)
- Goodfellow, I. Bengio, Y., Courville, A. (2016). Deep learning. MIT. (Ch.2) (есть на русском)
- Зельдович, Я.Б., Яглом, И.М. (1982) Высшая математика для начинающих физиков и техников. Наука. (Гл.2-5)