Введение. Аксиоматика Колмогорова.

Ссылка на видеокурс от МИАН

Введение

Стохастический анализ наиболее применимая область прикладной математики в **машинном обучении**. Практически в любом алгоритме машинного обучения используется **компонента стохастики**. Иными словами, в каком-нибудь алгоритме машинного обучения (МО) обязательно либо можно найти генератор псевдослучайных чисел, либо алгоритм будет запускаться на начальном наборе из случайных данных, либо по ходу будут выбираться ветки в зависимости от выпадения "орла" или "решки". Вариантов использования теории вероятностей, теории случайных процессов в машинном обучении много. Именно элемент стохастики дает вожможность надеяться, что эвристический метод МО даст результат, избегая переборного решения. Но *важно понимать*, что переборное решение дает *гарантированный ответ*, но на работу алгоритма уйдет огромное количество времени (ведь нужно перебрать все варианты), однако алгоритм МО эффективен, ответ на задачу будет получен в кратчайшие сроки, но *не гарантированно.* Стохастическая компонента – плата за то, что ответ может быть не получен.

К вопросу о случайности

Вопрос о том, что такое **случайность** довольно не прост. В нашем мире мы постоянно имеем дело со случайными явлениями, даже если такие явления не являются случайными в строгом смысле. В прикладных задачах это **не важно**.

Пример 1. Имеем три последовательности из нулей и единиц:

$$00000000000000\ldots \tag{1}$$

$$01010101010101\ldots \tag{2}$$

$$11011100111011\ldots \tag{3}$$

Встает вопрос: какие из этих последовательностей являются случайными? Кажется, что последовательности (1) и (2) не являются случайными: последовательность (1) – каждому элементу ставим в соответствие 0, последовательность (2) – каждому элементу ставим либо 0, либо 1 в зависимости от четности индекса элемента. Но что насчет (3)? Можно ли подобрать **алгоритм** и закодировать эту последовательность или же она случайная? На самом деле мы можем подобрать алгоритм. Число $\pi = 3,1415926535897\ldots$ – четной цифре ставим в соответствие 0, нечетной цифре – 1. То есть последовательность (3) не является случайной. Говорить о случайности имеет смысл лишь в применении к очень длинным цепочкам. Бессмысленно спрашивать, которая из цепочек 00, 01, 10, 11 более случайна, чем другие. Но у нас создается **впечатление случайности**, когда мы смотрим на последовательность (3) и не знаем, как она построена. Можно называть такую последовательность **псевдослучайной**.

Повторим еще раз важную вещь. Когда мы работаем со случайными объектами, вопрос о том, что случайно, а что нет, не имеет легкого разрешения, поэтому мы не будем его рассматривать. В качестве начального сета для алгоритма МО мы можем брать произвольный, удобный для нас набор данных, который может быть вовсе не случайным.

Вероятностное пространство. Аксиомы Колмогорова

Аксиоматический подход в теории вероятностей, развитый советским математиком Андреем Николаевичем Колмогоровым, является общепринятым.

Имеется некий бесконечно повторяемый случайный опыт. В результате одного опыта имеем некоторый случайный **элементарный исход** $\omega \in \Omega$.

Ω – некоторое множество из элементов любой природы, которые могут быть устроены **сколь угодно сложно**. Такое множество не может быть пустым. То есть сам ω может быть устроен сложным образом. К примеру, мы играем в "орлянку" и подбрасываем монетку бесконечно много раз — здесь ω фиксировано наперед – это результаты всех подбрасываний, которые мы можем увидеть.

Выберем подмножество $A \subset \Omega$. Говорят, что **событие** A произошло, если $\omega \in A$. К примеру, монету подбрасывают 15 раз (это есть Ω). Событие A состоит в том, что "решка" выпадает нечетное количество раз. В результате опыта имеем такой результат ω: "решка" выпала 9 раз, а "орел" – 6 раз. Таким образом, $\omega \in A$. На самом деле, если мы зафиксируем ω, то таких событий A, что $\omega \in A$, можно подобрать много, сделать это можно так, как нам удобно. Например, $A = \{$ "орел" выпадает четное количество раз $\}$ или $A = \{$ "решка" выпадает кратное трем число раз $\}$.

Важно понимать, что если в результате опыта мы видим событие A, то мы также видим событие $\overline{A} = \Omega \backslash A$ (дополнение к событию A).

Если A и B – события (например, в результате нашего опыта $A = \{$ "решка" выпала нечетное количество раз $\}$ и $B = \{$ "решка" выпала кратное трем число раз $\}$), то $A \cup B$ ("либо", например, "решка" выпала 5 раз, либо 6 раз) — событие, а также $A \cap B$ ("и", например, решка выпала 9 раз) — событие. Также мы потребуем, чтобы операции $\cup$ и $\cap$ были замкнуты относительно **счетного** числа событий. Например, если мы возьмем 5 событий, пересечем или объединим их, то получим новое событие.

Следует упомянуть, что само множество Ω – событие.

$\mathcal{A}$ – множество всех событий. Этот объект называется **σ-алгеброй** множеств над Ω. Вообще говоря, выбор σ-алгебры может быть произвольным, он не предусмотрен. Можно выбрать "богатую" σ-алгебру, можно ее упрощать. Разберемся на примере.

Пример 2. Допустим, что бросается шестигранный кубик. Тогда

$$\Omega = \{1, 2, 3, 4, 5, 6\}.$$

Если множество Ω конечно, то проще всего в качестве σ-алгебры взять множество всех подмножеств (все возможные комбинации из Ω). То есть мы рассматриваем все возможные события, которые можно получить в этом опыте. Но выбор "богатой" σ-алгебры не всегда удобен.

Рассмотрим такое множество всех событий:

$$\mathcal{A} = \{\Omega, \varnothing, \{1, 3, 5\}, \{2, 4, 6\}\}.$$

Это тоже σ-алгебра. Во-первых, в любой σ-алгебре присутствует Ω и (как дополнение) $\varnothing$. Во-вторых, выбор именно такой σ-алгебры позволяет заменить опыт в подбрасывании игральной кости на опыт в подбрасывании монетки (четное число точек – "орел", нечетное – "решка").

И, наконец, $\mathbb{P}$ – **вероятность**:

$$\mathbb{P} : \mathcal{A} \to [0, 1].$$

Вероятность – это функция, которая для каждого события выдает некоторое неотрицательное число, которое не превосходит 1. Если A – событие ($A \in \mathcal{A}$), то можно вычислить $\mathbb{P}(A) \in [0, 1]$. Для каждого события способ подсчета вероятности **одинаков**. На самом деле, строго говоря, $\mathbb{P}$ – вероятностная мера. Есть ряд технических моментов из теории меры, которые мы опустим.

Существуют дополнительные условия для $\mathbb{P}$:

- $\mathbb{P}(\Omega) = 1$
- $\mathbb{P}\left(\bigcup\limits_k A_k\right) = \sum\limits_k \mathbb{P}(A_k), A_m \cap A_l = \varnothing, m \neq l$

Если построена функция, удовлетворяющая всем вышеперечисленным условиям, то такая функция называется вероятностью.

Пара $\langle\Omega, \mathcal{A}\rangle$ называется **измеримым пространством**, вместе с $\mathbb{P}$ тройка $\langle\Omega, \mathcal{A}, \mathbb{P}\rangle$ называется **вероятностным пространством**. Если не построено вероятностное пространство, то мы не можем говорить о случайных событиях, вероятности и т.д., мы всегда должны фиксировать вероятностное пространство.

Случайная величина. Распределение вероятностей

Как мы уже говорили, $\omega \in \Omega$ могут быть сколь угодно сложными, поэтому над каждым элементом из Ω построим функцию для того, чтобы можно было сравнивать ω. Чтобы легче было сравнивать, функция будет принимать вещественные значения:

$$\xi : \Omega \to \mathbb{R}.$$

То есть, $\xi(\omega)$ – это некоторое вещественное число.

Пусть C – некоторая вещественная константа. Рассмотрим следующий вопрос:

$$\{\xi(\omega) < C\} = \{\omega \in \Omega : \xi(\omega) < C\}.$$

Этот вопрос фиксирует некоторое количество ω из Ω таких, что функция ξ от ω меньше, чем заданное число C. Мы задаем некоторое **событие**, поэтому $\{\xi(\omega) < C\} \in \mathcal{A}$. Соотвественно, функция ξ является **измеримой** относительно σ-алгебры $\mathcal{A}$. Мы должны фиксировать $\mathcal{A}$, относительно которой функция ξ является измеримой, потому что относительно одной $\mathcal{A}$ функция может быть измеримой, относительно другой – может быть не измеримой. Если рассматривать **пример 2**, то "богатая" σ-алгебра может ответить на вопрос, какое число выпало при броске кости (так мы задали функцию ξ), а другая σ-алгебра (упрощенная) не сможет дать ответ на этот вопрос (заданная ранее функция ξ не измерима относительно упрощенной σ-алгебры). В зависимости от того, какая задана σ-алгебра, определяется класс возможных функций ξ.

Класс таких функций ξ называется **случайная величина**. Случайным в данном случае является значение ξ, если случайным был аргумент ω, но сама функция ξ является строго детерминированной.

Зачастую, если интересна только случайная величина, то природа ω нас может вовсе не интересовать. Нас интересуют события, которые порождает заданная случайная величина.

Построим следующую числовую функцию (действующую из $\mathbb{R}$ в $\mathbb{R}$):

$$F_\xi(x) \stackrel{\text{def}}{=} \mathbb{P}\{\omega \in \Omega : \xi(\omega) < x\}.$$

- При $x \to -\infty$ – $F_\xi(x) \to 0$.
- При $x \to +\infty$ – $F_\xi(x) \to 1$.
- Эта функция неубывающая. С ростом x количество ω из Ω таких, что функция ξ от ω меньше, чем x, будет неуменьшаться.
- Эта функция непрерывна справа: $\lim\limits_{\varepsilon \to 0+} F_\xi(x+\varepsilon) = F_\xi(x)$.

Такую функцию называют **функцией распределения вероятностей**. Можно доказать, что для нее существует такое вероятностное пространство, которое будет ее реализовывать. Поэтому необязательно его фиксировать. Имея функцию распределения, мы можем оперировать случайными величинами.

Если функция F_ξ окажется дифференцируемой

$$F'_\xi(x) = f_\xi(x),$$

то получаем f_ξ – функцию **плотности распределения**.

- Поскольку F_ξ – неубывающая, то f_ξ – неотрицательная.

- $\int\limits_{-\infty}^{\infty} f_\xi(x)\,dx = 1.$

Дальше возникает вопрос о **независимости (несовместности)** двух случайных величин ξ и η. Чтобы дать на него ответ, нужно посчитать **совместную плотность распределения**.

Рассмотрим две компоненты случайного исхода $\omega = (\omega_1, \omega_2)$. Пусть одна случайная величина ξ определяется ω_1, другая, η – от ω_2. Если в опыте понятно, что w_1 не зависит от w_2 (например, одновременно подбрасываются 2 монеты), то случайные величины ξ и η независимы. Чтобы точно ответить на вопрос о несовместности, нужно посчитать совместную плотность распределения. Чтобы ее сосчитать, необходимо вернуться к вероятностному пространству. Таким образом, вопрос о независимости случайных величин сводится к вопросу о независимости **случайных событий**.

Два события A и B являются независимыми, если $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$.

Пример 3. Подбрасывается честный игральный кубик:

$$\Omega = \{1, 2, 3, 4, 5, 6\}.$$

Рассмотрим $\mathcal{A}$, состоящее из всех подмножеств Ω. Пусть событие $A = \{1, 2, 3\} \in \mathcal{A}$, событие $B = \{3, 4\} \in \mathcal{A}$. Сосчитаем их вероятность. Вероятность события A – $\mathbb{P}\{A\} = \frac{3}{6} = 0.5$, вероятность события B – $\mathbb{P}\{B\} = \frac{2}{6} = \frac{1}{3}$. Вероятность их пересечения – $\mathbb{P}\{A \cap B\} = \mathbb{P}\{\{3\}\} = \frac{1}{6}$. Выполняется равенство $\mathbb{P}\{A \cap B\} = \mathbb{P}\{A\} \cdot \mathbb{P}\{B\}$, значит события A и B – независимы.

Мы наглядно показали, что стохастическая система полностью строится над вероятностным пространством.