

Основные понятия и определения

Шокуров Антон В.
shokurov.anton.v@yandex.ru
<http://машинноезрение.рф>

2 марта 2017 г.

Версия: 0.10

Аннотация

Вводятся базовые, традиционные понятия, а точнее принятый в этой области жаргон и его математический эквивалент. В дальнейшем сам жаргон не планируется применяться.

Предварительная версия!

1 Введение

В данной области информатики изучается задача построения прогноза некоего явления по неким наперед зафиксированным данным, которые по постановке считаются непосредственно связанными с прогнозируемым явлением. Метод построения прогноза основан на изучении “исторических” полных данных, состоящих как из самих данных, так и правильного результата (явления). С данными необязательно связано время, поэтому историческими данными могут быть просто некие прецеденты.

Пример 1. Прогноз погоды. Хотим дать прогноз погоды. В самом общем случае никто нас не ограничивает инструментарием.

Пример 2. Фондовый рынок. Актуальной задачей является прогноз цены ценных бумаг. Необходимо по истории поведения цены ценных бумаг спрогнозировать их стоимость в будущем. В такой постановке предполагается, что изменение цены в будущем определяется в том числе и на основе исторических данных. На самом деле отмеченное предложение уже фактически является частью решения заявленной задачи. Ясно, что например, открытие нового месторождения ископаемых или аварию на шахте мы не прогнозируем (хотя они окажут

резко изменение в цене акций). Но, даже и в таком случае, исторические данные позволяют понять, что в общем можно ожидать.

Последний пример может создать впечатление, что для прогноза могут быть использованы только данные связанные с котировками ценных бумаг. На самом деле, для прогноза может пригодиться, как ни странно, и решение предыдущего примера. Суть в том, что есть такая вещь как Ла-Нинья/Эль-Ниньо, от которой многое чего зависит. Например, сила ливней и соответственно факт затопления шахт в Австралии. Когда они затоплены цены на уголь обычно уходят в космос. Исторические котировки угля конечно могут схожий результат проследить, но они не смогут предсказать очередное Эль-Ниньо.

Пример 3. Студенты. ФИО. *В качестве контраста приведем не совсем правильный пример. Допустим нужно (преподавателю) запомнить студентов группы. Он не правильный потому что прогноза не будет. Достаточно будет только запомнить.*

Если последнюю задачу чуть переформулировать, то это уже будет задачей машинного обучения.

Пример 4. Студенты. Оценка *Нужно спрогнозировать оценку которую получит студент.*

1.1 Постановка задачи

Перейдем к математическому языку. Задано множество объектов Ω и множество значений Γ , а также искомого отображения $\gamma : \Omega \rightarrow \Gamma$. Следует сразу отметить неоднозначность выбора множеств Ω и Γ даже в зафиксированной и обговоренной задаче (на подобие ранее приведенных примеров). При их фиксации в виде конкретных множеств задача уже будет носить частный характер (т.е. выбор этих множеств уже является неким этапом в решении задачи).

Отметим, что в общем случае на Ω никаких тождеств не задано и не выполнено. Например, в теории вероятности требуют выполнения соответствующих аксиом. В рассматриваемом случае никаких таких дополнительных требований нет. Последнее означает, что Ω есть суть множество элементов ω .

Пример 5. Прогноз погоды. *Например, хотим по некий данным дать предсказание погоды на некий день. Много вариантов. Можно спрогнозировать на разные периоды: день, неделя, месяц и тому подобное. Можно спрогнозировать разные явления: температуру, осадки и тому подобное. Можно спрогнозировать среднюю температуру, а можно максимум.*

Пример 6. Фондовый рынок. В качестве Ω можно взять исторические котировки (конкретного эмитента), а можно рассматривать и исторические события (отчетности, ключевые новости). Множество Γ также может принимать широкий диапазон: как прогноз конкретной цены на завтра (процентная величина изменения от текущей цены), или просто утверждение о том, что цена не уйдет ниже/выше какого-то значения.

Пусть $\Omega^l \subseteq \Omega$. По построению, для $\forall \omega \in \Omega^l$ можно вычислить значение $\gamma(\omega)$. Следовательно, по Ω^l можно построить множество пар

$$\Lambda^l = \{(w, \gamma(w)) | w \in \Omega^l\}.$$

Тогда, задачу можно сформулировать следующим образом: необходимо по Λ^l построить функцию $\tilde{\gamma}$ приближающую искомую функцию γ , т. е. для $\forall \omega \in \Omega$ вычислить приближение к $\gamma(\omega)$. Подчеркну, что саму γ использовать для построения нельзя. Она может быть использована только для оценки итогового качества.

В точки зрения науки *машинного обучения* традиционно произносится следующее словоблудие (жаргоны): Ω^l – множество произвольно *выбранных обучающих* элементов $\omega \in \Omega$, где l от слова learn (т.е. *обучение*). Процесс построения функции $\tilde{\gamma}$ называется *обучением*. Пара $(\omega, \gamma(\omega))$, где $\omega \in \Omega^l$ – суть *прецедент*, а множество пар Λ^l – *обучающая выборка*, т. е. то, что нам фактически дано для *обучения*. Последнее предполагает возможность *обобщения*, т.е. вычисления значения отображения $\tilde{\gamma}$ в произвольной точке Ω (а не только на самом множестве Ω^l).

С практической точки зрения, естественно, часто предполагается, что введенные множества конечны, а именно – $\Omega^l = \{\omega_i\}_{i=1}^l$, $\Lambda^l = \{(\omega_i, \gamma_i = \gamma(\omega_i))\}_{i=1}^l$.

Пример 7. Фондовый рынок. Данные можно выбирать на *тиковые*, а можно *минутные*, а можно и *дневные*. В зависимости от данного выбора будет зависеть и конкретная частная постановка задачи.

Векторное пространство Для начала рассмотрим крайний случай, а именно – наложим на множество Ω выполнение аксиом векторного пространства. Последнее означает, что на множестве Ω дополнительно (в общем случае никаких операций на Ω не заданы) задана две операции $+$ и $*$. Напомним аксиомы....

Тогда оно есть суть векторное пространства \mathbb{R}^n размерности n .

Пусть также искомая функция γ линейна (опять же в общем случае таких требований не накладывается). Тогда легко восстановить функцию по линейно независимым прецедентам ($\{e_i\}_{i=1}^n$). Так, из линейной независимости $\{e_i\}_{i=1}^n$ следует, что для $\forall \omega \in \Omega = \mathbb{R}^n$, $x = \omega$ верно $x = \sum_1^n x_i e_i$, следовательно

$$\gamma(x) = \gamma\left(\sum_1^n x_i e_i\right) = \sum_1^n x_i \gamma(e_i).$$

Таким образом значение на произвольном x определяется значением на базисных элементах (прецедентах).

В общем случае это конечно не так, поэтому приблизимся к такому случаю в следующем подразделе.

1.2 Типы множеств

В общем случае про Ω ничего не известно, в частности, на нем не определены вычислительные операции (как это было в ранее рассмотренном примере векторного пространства). Но, ввиду того, что данный курс предполагает построение некой функции в смысле математики (а также тезиса, что „все есть число“), множества Ω и Γ предполагают некую математическую доработку или, иначе, подкрутку.

Функция от одного аргумента По аналогии с случайными величинами в теории вероятности, пусть существуют функции $\{\varphi_j : \Omega \rightarrow \Psi\}_{j=1}^{j=n}$. Фактически функция φ_j по сути соответствует искомой задаче, т.е. мы сами строим n вспомогательных функций $\{\varphi_j\}_{j=1}^{j=n}$ схожих функций на γ . Разница в том, что в отличие от γ мы знаем устройство функций φ_j . Смысл этих функций в том, чтобы представить искомую γ через них.

Дабы это все имело смысл про Ψ уже требуется, что оно является алгебраическим множеством, т. е. фактически множества наделенные операциями $+$, $-$, $*$, $/$, а иногда и порядковыми $<$, $>$. Наиболее традиционными являются числовые множества: \mathbb{R} , \mathbb{Z} , \mathbb{Z}_2 , \mathbb{N} и тому подобные. Разумеется, что в качестве такого множества может выступить и \mathbb{R}^k : $\varphi \rightarrow \mathbb{R}^k$. В последнем случае, если не обговорено иное, заменим такую функцию φ на множество функций $\{\varphi_j : \Omega \rightarrow \mathbb{R}\}_{j=1}^{j=k}$, перенумеруем их и включим в общий список.

Для множества Γ по аналогии строиться функция $\psi : \Gamma \rightarrow \Psi$, где от ψ дополнительно требуют существование обратной функции. Если не обговорено обратного, всюду далее считаем, что Ψ есть суть \mathbb{R} .

Тогда, фактически, задача поиска значения искомой функции γ в ω будет заменена на приближение значения $\psi(\gamma(\omega))$ по значениям $\varphi_i(\omega)$.

С точки зрения науки *машинного обучения* данный этап один из самых важных при решении конкретной задачи, в частности потому, что считается, что он делается так называемым *экспертом*. От правильности выбора (им) этих функций (которые на жаргоне называются *признаками*) фактически зависит успех в решении задачи.

Пример 8. Фондовый рынок. В качестве признаков можно рассматривать цену в какой-то момент времени, а можно и значение цены за день, час, минуту. От выбора будет зависеть успех выбранного метода. Можно также минимум/максимум, открытие/закрытие.

Так как считается, что функции φ_i и ψ задаются экспертом, то считается что они зафиксированы. Последнее означает, что их построение как бы не является предметом рассмотрения науки машинного обучения. Считается, что этих значений достаточно для решения поставленной задачи.

Ввиду введенных ограничений часть ранее введенных определений может быть переформулирована следующим образом. Произвольному $\omega \in \Omega$ можно сопоставить вектор-строку

$$\varphi(\omega) = (\varphi_1(\omega), \dots, \varphi_n(\omega)) = (x_1, \dots, x_n) = x,$$

называемый на жаргоне *признаковым описанием* объекта ω . Никто не говорит, что появится смысл в сложении в признаковом пространстве.

Тогда функцию $\tilde{\gamma}$, которую мы ищем, можно представить в виде: $\tilde{\gamma} = \psi^{-1}(\tilde{f}(\varphi))$, где, например,

$$\tilde{f} : \mathbb{R}^n \rightarrow \mathbb{R}.$$

Если зафиксировать функции φ_i и ψ , то первоначальная задача сводится к построению функции \tilde{f} , которая зависит от математически понятных величин.

Уравнение также можно переписать в виде $\psi(\tilde{\gamma}) = f(\varphi)$. Пусть $y_i = \psi(\gamma(\omega_i))$, где $i = 1, \dots, i = l$.

Можно задавать и неявной функцией, но зачем усложнять...

Отметим, что может сложиться следующая ситуация: $\exists \omega_a, \omega_b \in \Omega$, $\omega_a \neq \omega_b$ такие что $\varphi(\omega_a) = \varphi(\omega_b)$, т.е. признаковые описания могут совпадать у не совпадающих элементов. Иначе говоря, в общем случае, функция φ необратима (отметим требование обратимости ψ).

Более того, мы будем предполагать, что на самом деле каждая из φ не обязательно определена на всем множестве Ω , т.е. может $\exists \omega \in \Omega$ такое, что значение на каких-то из φ не определено.

Тем самым по Ω^l можно построить числовую матрицу данных

$$X = \begin{pmatrix} x_1^1 \cdots x_n^1 \\ \vdots \\ x_1^i \cdots x_n^i \\ \vdots \\ x_1^l \cdots x_n^l \end{pmatrix} = \begin{pmatrix} x_1 \\ \vdots \\ x_i \\ \vdots \\ x_l \end{pmatrix} = \begin{pmatrix} \varphi(\omega_1) \\ \vdots \\ \varphi(\omega_i) \\ \vdots \\ \varphi(\omega_l) \end{pmatrix} = \begin{pmatrix} \varphi_1(\omega_1) \cdots \varphi_n(\omega_1) \\ \vdots \\ \varphi_1(\omega_i) \cdots \varphi_n(\omega_i) \\ \vdots \\ \varphi_1(\omega_l) \cdots \varphi_n(\omega_l) \end{pmatrix}$$

и соответственно матрицу значений

$$Y = \begin{pmatrix} y_1 \\ \vdots \\ y_i \\ \vdots \\ y_l \end{pmatrix} = \begin{pmatrix} \psi(\gamma(\omega_1)) \\ \vdots \\ \psi(\gamma(\omega_i)) \\ \vdots \\ \psi(\gamma(\omega_l)) \end{pmatrix}.$$

Не все значения Y могут существовать.

У нас есть матрица X и вектор (значений) Y . Можно составить систему линейных уравнений: $X\theta = Y$, где $\theta = (\theta_1, \dots, \theta_n)$.

Задание. Написать программу которая решает заданную систему в среднеквадратичном смысле. Применить к поиску прямой проходящей через точки лежащие почти на прямой: берем множество точек равномерно распределенных на отрезке $[a, b]$, далее для каждой точки строим значение y , далее добавляем к каждой из точек y на нормальное распределение. Ищем прямую проходящую через эти точки наилучшим образом в среднеквадратичном смысле.

Задание. Все аналогично предыдущему случаю. Но вместо прямой рассмотрим параболу, окружность, эллипс.

Функция от двух аргументов По аналогии экспертом можно ввести функцию и от двух аргументов. Это будет что-то наподобие метрики, т.е. она будет вычислять расстояние между объектами. Тривиальный способ применить метрику в действительных векторных пространствах. $\rho(\omega_0, \omega_1) = \|\varphi(\omega_0) - \varphi(\omega_1)\|$. Таким образом такую метрику всегда можно задать, но лучше если её построит эксперт исходя из частных соображений об объектах в Ω .

В некоторых из методов обучения мы применим данную функцию. Смысл в не в том, что отсутствует переход в векторное пространство. Функции с большей арности обычно не применяются.

Иерархия признаков пространств Естественно, что к X и Y можно отнести как к $\hat{\Omega}$ и $\hat{\Gamma}$. Например, в новом признаковом пространстве будет добавлено произведение двух каких-то признаков. Тогда, всю иерархию промежуточных признаков пространств можно выкинуть.

1.3 Класс задачи

В информатике редко когда удается эффективно/рационально решить задачу в общем случае. Поэтому и в машинном обучении многие методы построены на неких

особенностях/частностях формулировки задачи, и поэтому принято выделять/обослаблять определенные типы множеств, относящиеся к определенным типам задач.

Область значения в большей степени определяет класс задачи. Если множество бесконечное, непрерывное, то это задача регрессии (есть понятие близости, можно ошибиться на чуть-чуть.). Если оно конечное, то это классификация/кластеризация (нет понятия близости. ошибиться чуть чуть нельзя.).

Пример 9. *Пример о распознавании лиц на изображениях...описание сцены...*

Физика... трение и притяжение...

Животные ... классовый вид...

Кредит...

Разные типы ошибок: аппроксимация какой-то кривой и проведение кривой приближающей данные.

1.4 Качество восстановления

Функцию можно по разному восстанавливать (т.е. обучать) или продолжать. Необходим способ оценки качества восстановления. В случае фиксации способа оценки качества, решение задачи сводится к поиску функции минимизирующая данную оценку. Последнее означает, что от выбранного способа оценки будет существенно зависеть и метод решения задачи.

В единственной точке Пусть у нас есть две функции кандидата f_0 и f_1 на приближение искомой функции. Нам нужно понять какая функция лучше приближает. Мы можем проверить какое эти функции дают значение в некой точке $x_0 = \varphi(\omega_0)$, $\omega_0 \in \Omega$ и сравнить его с истинным $y_0 = \psi(\gamma(\omega_0))$.

В рамках данного обсуждения введем функцию близости:

$$d : \Gamma \times \Gamma \rightarrow R_+$$

измеряющая схожесть между двумя значениями из Γ . Она также может быть задана и на Ψ

$$d : \Psi \times \Psi \rightarrow R_+.$$

Будем считать что у нас есть полиморфизм. Также учтем, что зная одно можно построить другое. Так, $\forall \psi_0, \psi_1$ можно положить $d(\psi_0, \psi_1) = d(\psi(\psi_0), \psi(\psi_1))$. И наоборот, $\forall y_0, y_1$ можно положить $d(y_0, y_1) = d(\psi^{-1}(y_0), \psi^{-1}(y_1))$.

Заметим, что это функция d не тоже самое, что функция ρ близости задаваемая в Ω .

В качестве d обычно рассматривают:

Пример 10. *Индикатор ошибки:* $d(u, v) = (u \neq v)$. Фактически является булевским значением. При численных вычислениях может быть представлено либо: 0 и 1 либо: -1 и 1. В зависимости от контекста.

Пример 11. *Величина отклонения:* $d(u, v) = |u - v|$. Является модулем расстояния между точками.

Пример 12. *Квадрат модуля отклонения:* $d(u, v) = (|u - v|)^2$.

Пример 13. *Произвольная степень p модуля ошибки:* $d(u, v) = (|u - v|)^p$.

Этими расстояниями список не оканчивается. Вообще её можно задавать произвольно. От выбора этой функции зависит успешность численного решения задачи. Например, в некоторых случаях квадратичная метрика даст точное решение, тогда как другие метрики потребуют определенного перебора (итеративный алгоритм), например, последовательное приближение.

Множество точек В общем случае качество восстановления функции хотелось бы проверить на множестве значений, а не в одной точке. Последнее связано с тем, что в реальных задачах важнее то, как функция ведет себя на множестве точек.

Традиционно сравнивают функции в среднем, но можно и экстремальные значения (максимум/минимум) проверить.

Пример 14. Прогноз погоды. При прогнозе погоды часто бывает, что важна именно средняя температура, а не температура в конкретный день. Точнее даже важны такие вещи как возможность засухи или болота.

Пример 15. Фондовый рынок. Для фондового рынка важно, что в среднем прибыль есть, а не то, что она в любой сделке имеется.

Интегральная оценка качества задается следующим образом. Сначала задается множество $\Omega^t \subseteq \Omega$ элементов на которых и будет оценена построенная функция f .

$$\Lambda(\tilde{f}, f, \Omega^t) = \frac{1}{|\Omega^t|} \sum_{\omega \in \Omega^t} d(f(\varphi(\omega_i)), \tilde{f}(\varphi(\omega_i))).$$

Замечание. Можно формулу объяснить (усложнить) и посредством вероятностного подхода. Так, появится плотность распределения в качестве множителя. Но мы пока не будем усложнять формулу таким образом.

Опять же из практических соображений обычно отмеченное множество конечно: $\Omega^t = \{\omega_i^t\}_{i=1}^l, \omega_i \in \Omega$.

тестовых Можно было бы и взять максимум/минимум.

Вопрос выбора тестового множества будет рассмотрен отдельно.

1.5 Обучение

Задается экспертом. Берется некое параметрическое множество функций $\{g : \Omega \times \Theta \rightarrow \Upsilon\}$, т.е. функции вида $g_\theta(\omega) = g(\omega, \theta)$. Суть обучения заключается в *оптимальном* выборе значения θ . Данное семейство задается экспертом и от выбора этого множества существенным образом зависит успех решения задачи.

Гиперпараметр настраивается тестовым множеством.

Обобщающая способность! Не должна быть памятью. Переобучение и недообучение.

Не всему можно обучиться: Что-то с данными: 0) Плохо выбраны признаки.
1) Недостаточно данных. 2) На уровне x_i и y_i много шума. 3) Неоднозначный ответ.

Что-то с методов обучения. Он не подходит для данной задачи.