Методы машинного обучения

Лекция 5

Классификация - Байесовский подход

Байесовский подход

Основные определения

- Совместная вероятность – вероятность одновременного наступления двух событий: $P(A, B)$;
- Независимость: A и B независимы, если: $P(A, B) = P(A)P(B)$
- Условная вероятность – вероятность наступления одного события, если известно, что произошло другое, $P(A|B)$ - вероятность наступления события A при условии, что событие B произошло. Очевидный частный случай: P(A|A)=1=100%
- Вероятность совместного появления двух зависимых событий равна:

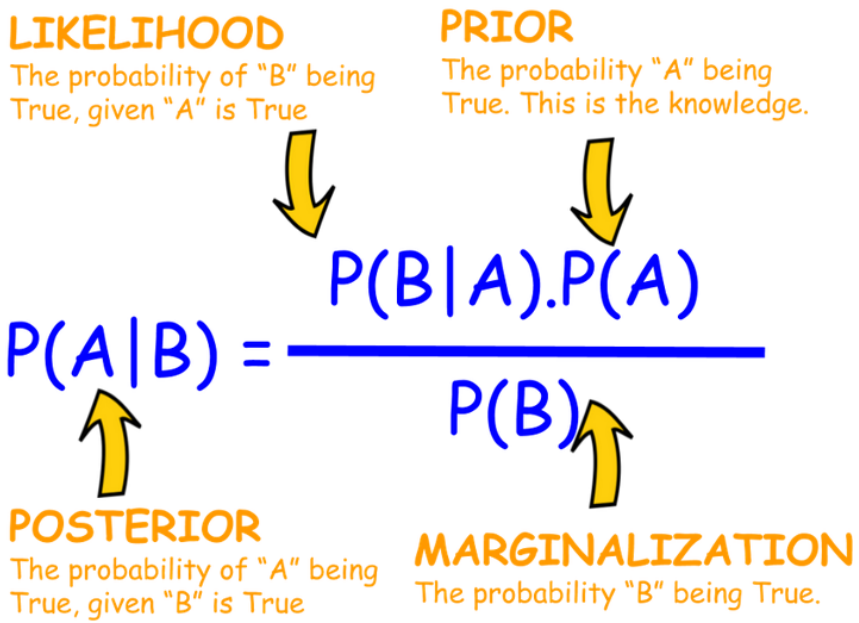
$$P(A, B) = P(A|B)P(B) = P(B|A)P(A)$$

- Теорема Байеса – из предыдущей формулы:

$$P(A|B) = \frac{P(A, B)}{P(B)} = \frac{P(B|A)P(A)}{P(B)}$$

Теорема Байеса

Теорема Байеса (или формула Байеса) — одна из основных теорем элементарной теории вероятностей, которая позволяет определить вероятность какого-либо события при условии, что произошло другое статистически взаимозависимое с ним событие. Другими словами, по формуле Байеса можно более точно пересчитать вероятность, взяв в расчёт как ранее известную информацию, так и данные новых наблюдений.



- P(A) — априорная вероятность гипотезы A (prior probability);
- P(A|B) — вероятность гипотезы A при наступлении события B (апостериорная вероятность - posterior probability);
- P(B|A) — вероятность наступления события B при истинности гипотезы A (правдоподобие - likelihood);
- P(B) — полная вероятность наступления события B (вероятность данных evidence).

Вычисление $P(B)$ - Маргинализация

В задачах и статистических приложениях $P(B)$ обычно вычисляется по формуле полной вероятности события, зависящего от нескольких несовместных гипотез, имеющих суммарную вероятность 1.

$$P(B) = \sum_{A} P(A, B) = \sum_{i=1}^{N} P(B|A_i)P(A_i)$$

где вероятности под знаком суммы известны или допускают экспериментальную оценку.

Тогда **формула Байеса** примет вид:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{P(B|A)P(A)}{\sum_{i=1}^{N} P(B|A_i)P(A_i)}$$

Пример: О болезнях и вероятностях

Пусть некий тест на какую-нибудь болезнь имеет вероятность успеха 95% (т.е. 5% - вероятность как позитивной, так и негативной ошибки). Всего болезнь встречается у 1% респондентов (не учитываем, что они разного возраста и профессий).

Пусть некий человек на диспансеризации неожиданно для него получил позитивный результат теста (тест говорит, что у него есть заболевание). С какой вероятностью он действительно болен?

Вывод с использованием байесовского подхода

Обозначим через t результат теста, через d-наличие болезни.

Тогда:

$$p(t=1) = p(t=1|d=1)p(d=1) + p(t=1|d=0)p(d=0)$$

Используем теорему Байеса:

$$p(d=1|t=1) = \frac{p(t=1|d=1)p(d=1)}{p(t=1|d=1)p(d=1) + p(t=1|d=0)p(d=0)}$$

$$= \frac{0{,}95 \times 0{,}01}{0{,}95 \times 0{,}01 + 0{,}05 \times 0{,}99} = 0{,}16$$

Выводы:

Такие задачи составляют суть вероятностного вывода (probabilistic inference). Поскольку они обычно основаны на теореме Байеса, вывод часто называют байесовским (Bayesian inference).

Интерпретация вероятности

Вероятность как частота

- Обычно в классической теории вероятностей, происходящей из физики, вероятность понимается как предел отношения количества определенного результата эксперимента к общему количеству экспериментов.
- Стандартный пример: бросание монетки.

Вероятность, как степень доверия

Мы можем рассуждать о том, «насколько вероятно» то, что:

- Динозавры вымерли в результате падения метеорита;
- Произойдет столкновение с определенным космическим объектом;
- В соседних звездных системах есть жизнь;
- Сборная России победит на ближайшем чемпионате мира по футболу.

Говорить о «стремящемся к бесконечности количестве экспериментов» совершенно бессмысленно, т.к. эксперимент здесь ровно один. Вероятности выступают как степень доверия (degrees of belief). Это байесовский подход к вероятностям (Томас Байес так их понимал).

Прямые и обратные задачи

Прямая задача:

- В корзине лежат 10 шаров, из них 3 черных. Какова вероятность выбрать черный шар?
- В корзине лежат 10 шаров с номерами от 1 до 10. Какова вероятность того, что номера трех последовательно выбранных шаров дадут в сумме 12?

Обратная задача:

- Перед нами две корзины, в каждой по 10 шаров, но в одной 3 черных, а в другой 6. Некто взял из какой-то корзины шар, и он оказался черным. Насколько вероятно, что он брал шар из первой урны?

В обратной задаче вероятности сразу стали байесовскими, т.е. необходимо определить вероятность какого-либо события при условии, что произошло другое статистически взаимозависимое с ним событие.

Иначе говоря, прямые задачи теории вероятностей описывают некий вероятностный процесс или модель и просят подсчитать ту или иную вероятность (т.е. фактически по модели предсказать поведение). Обратные задачи содержат скрытые переменные (в примере – это номер корзины, из которой брали шар) и часто просят по известному поведению построить вероятностную модель.

Теорема Байеса позволяет переставить местами причину и следствие. Зная с какой вероятностью причина приводит к некоему событию, эта теорема позволяет рассчитать вероятность того что именно эта причина привела к наблюдаемому событию.

ML(maximum likelihood) vs. MAP (maximum a posteriori)

В статистике обычно ищут гипотезу максимального правдоподобия (maximum likelihood):

$$\theta_{ML} = arg\max_{\theta} p(D|\theta)$$

В байесовском подходе ищут апостериорное распределение (posterior)

$$p(\theta|D) \propto p(D|\theta)p(\theta)$$

и, возможно, максимальную апостериорную гипотезу (maximum a posteriori):

$$\theta_{MAP} = arg\max_{\theta} p(\theta|D) = arg\max_{\theta} p(D|\theta)p(\theta)$$

Функция правдоподобия имеет вид:

$$a \mapsto p(y|x = a)$$

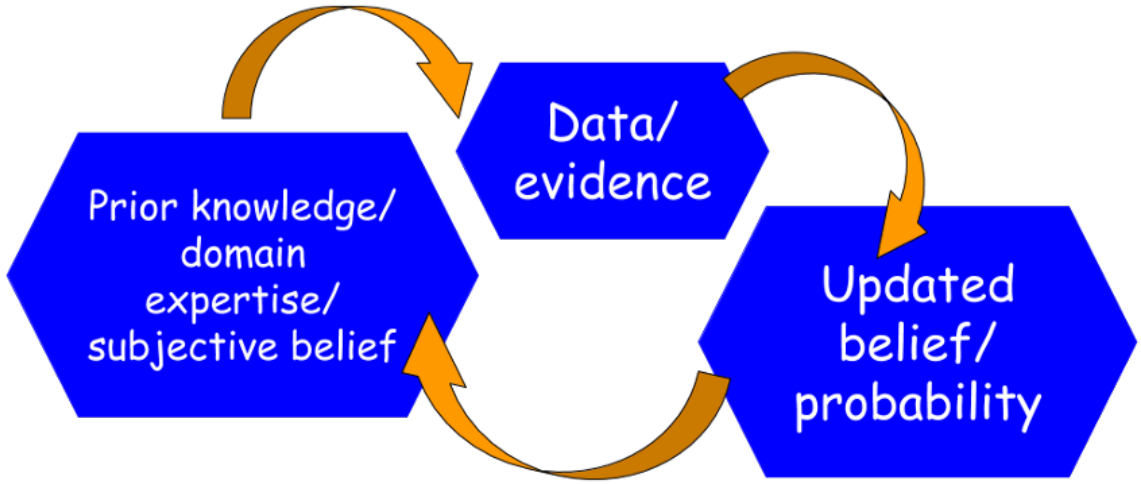
Логический процесс анализа данных

Исторически в большинстве методов статистического обучения (статистических исследований) понятие априорного события не используется или недооценивается.

Теорема Байеса позволяет учитывать субъективную оценку или уровень доверия в строгих статистических расчетах. Это один из методов, который позволяет постепенно обновлять вероятность события по мере поступления новых наблюдений или сведений.

- Мы начинаем с гипотезы и уровня доверия к этой гипотезе. Это означает, что на основе знания предметной области или предшествующих других знаний мы приписываем этой гипотезе ненулевую вероятность.
- Затем мы собираем данные и обновляем наши первоначальные убеждения. Если новые данные подтверждают гипотезу, то вероятность возрастает, если не подтверждают - вероятность снижается.

Звучит просто и логично, неправда ли?



Пример: Скрининг-тест на употребление наркотиков

Предположим, что тест на применение наркотика имеет **97% чувствительность** (доля истинно положительных результатов) и **95% специфичность** (доля истинно отрицательных результатов). То есть тест даст **97% истинно положительных** результатов для потребителей наркотиков и **95% истинно отрицательных результатов** для лиц, не употребляющих наркотики. Предположим, мы также знаем, что 0,5% населения в целом употребляют наркотики.

Какова вероятность того, что случайно выбранный человек с положительным результатом анализа является потребителем наркотиков?

Знание о проценте употребляющих является важнейшей частью «**априорной вероятности**», которая представляет собой часть обобщенных знаний об общем уровне распространенности. Это наше предварительное суждение о вероятности того, что случайный испытуемый будет употреблять наркотики. Это означает, что **если мы выберем случайного человека из общей популяции без какого-либо тестирования, мы можем только сказать, что вероятность того, что этот человек употребляет наркотики, составляет 0,5%.**

Расчет по правилу Байеса

Формула для вычисления по правилу Байеса:

- принимает в качестве входных данных чувствительность и специфичность теста, а также предварительные знания о проценте потребителей наркотиков в популяции;
- выдает вероятность того, что тестируемый является потребителем наркотиков, на основе положительного результата теста.

$$P(U^+|T^+) = \frac{P(T^+|U^+)P(U^+)}{P(T^+)} = \frac{P(T^+|U^+)P(U^+)}{P(T^+|U^+)P(U^+) + P(T^+|U^-)P(U^-)}$$

$P(U^+)$ = Уровень распространенности наркомании

$P(U^-)$ = 1 - Уровень распространенности наркомании

$P(T^+|U^+)$ = Чувствительность теста

$P(T^-|U^-)$ = Специфичность теста

$P(T^+|U^-)$ = 1 - Специфичность теста

$$P(U^+|T^+) = \frac{0{,}97 \cdot 0{,}005}{0{,}97 \cdot 0{,}005 + 0{,}05 \cdot 0{,}995} = 0{,}089$$

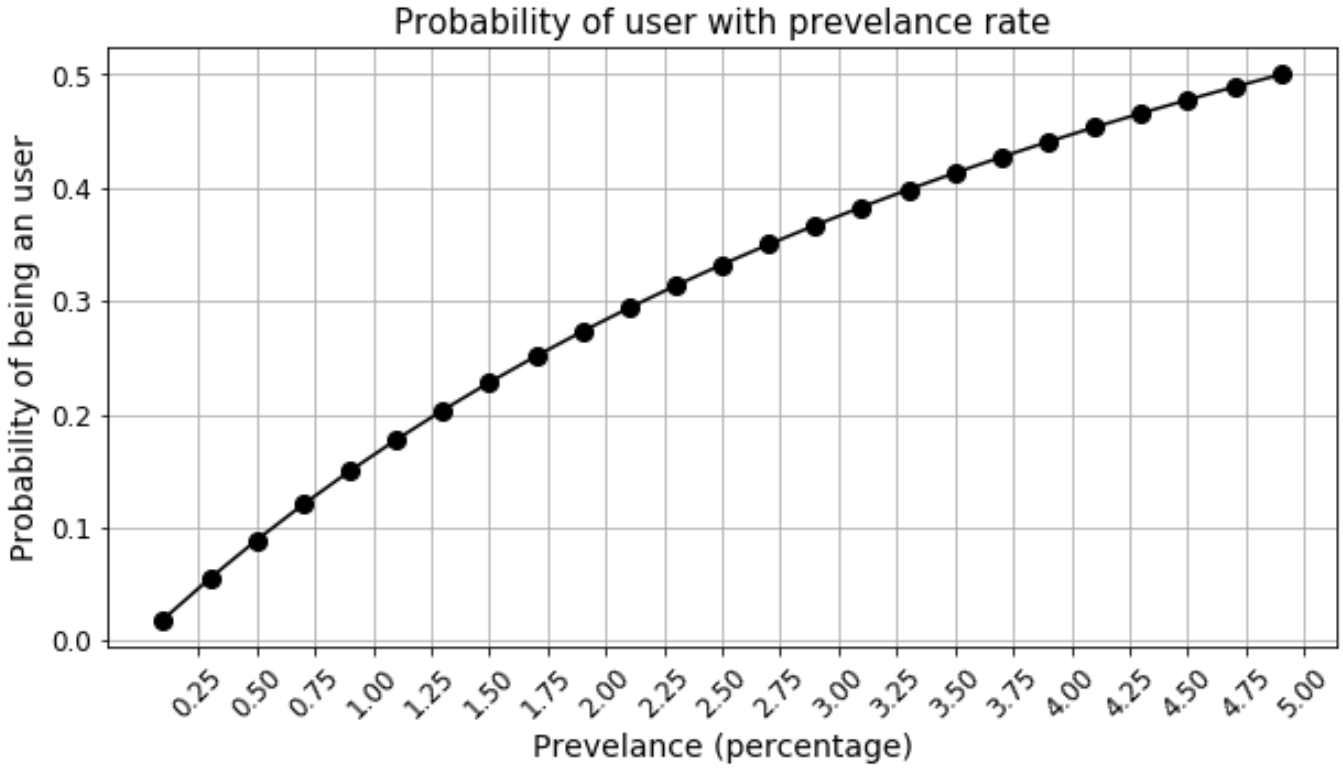
Что здесь интересного?

Даже при использовании теста, который в 97% случаях верно выявляет положительные случаи и который в 95% случаях правильно выявляет отрицательные случаи, истинная вероятность выявить человека употребляющего наркотики с положительным результатом теста составляет всего 8,9%!

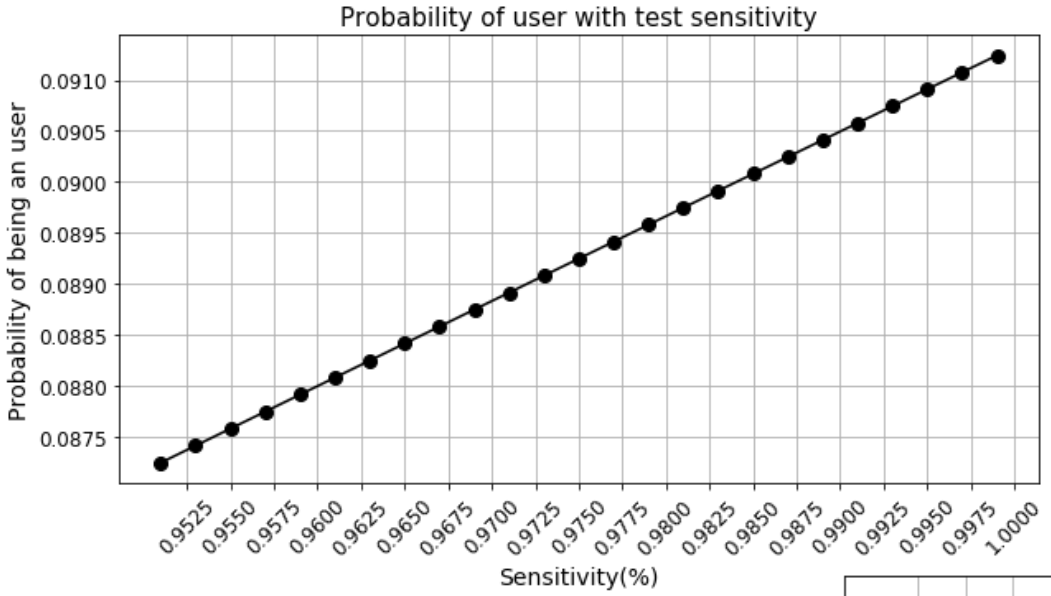
Всего 8,9%! Что не так?

Это связано с чрезвычайно низким уровнем распространенности. **Количество ложных срабатываний превышает количество истинных срабатываний.**

Если протестировано 1000 человек, ожидается, что будет 995 не наркоманов и 5 наркоманов. Из 995 не наркоманов ожидается 0,05 × 995 ≃ 50 ложных срабатываний. Из 5 наркоманов ожидается 0,97 × 5 ≈ 5 истинно положительных результатов. Из 55 положительных результатов только 5 являются истинно положительными!

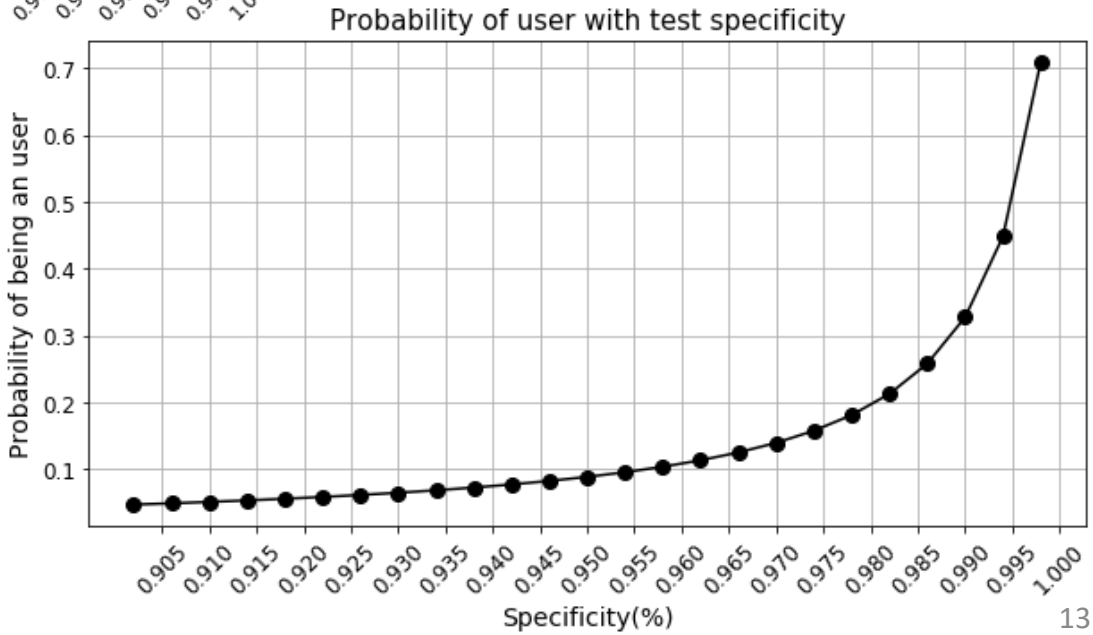


Какой уровень точности теста необходим для улучшения сценария?



Даже с чувствительностью, близкой к 100%, улучшения практически не происходит.

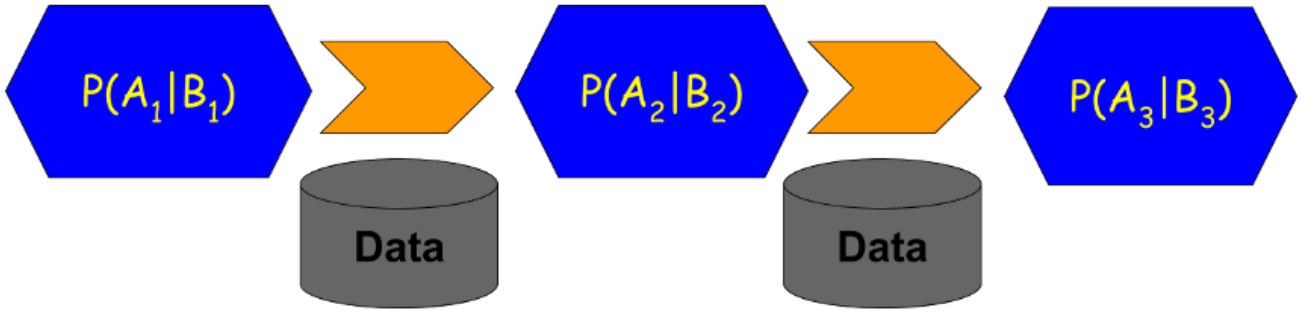
Имеет место нелинейная зависимость вероятности от специфичности теста, и по мере увеличения специфичности, происходит значительное увеличение вероятности.



Цепочка расчетов и формула Байеса

Лучшее в байесовском выводе - это **возможность использовать предшествующие знания** в форме **априорного** вероятностного члена в числителе теоремы Байеса.

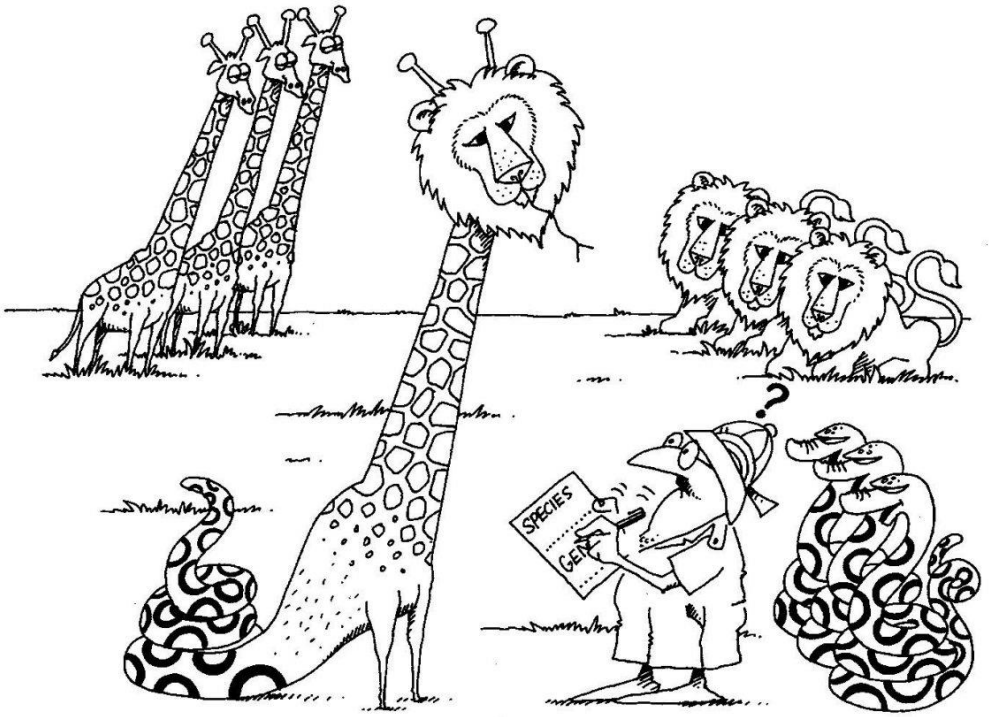
- Предварительные знания - это не что иное, как вычисленная вероятность теста, которая затем возвращается к следующему тесту.
- Для случаев, когда уровень распространенности среди населения в целом чрезвычайно низок, один из способов повысить уверенность в результате теста - назначить последующий тест, если первый результат теста окажется положительным.
- *Апостериорная* вероятность первого теста становится *априорной вероятностью* для второго теста, т.е. $P(U^+)$ для второго теста уже не общий показатель распространенности, а вероятность из первого теста.
- Расчетная (апостериорная) вероятность первого теста 8,9%, во втором тесте она возрастает до 65,4%, а третий положительный тест дает значение 97,3%.
- Следовательно, неточный тест можно использовать несколько раз, чтобы обновить мнение с помощью последовательного применения правила Байеса.



Chaining of Bayes' rule for updating probabilities as the data comes in

Классификация

Классификация (classification) — это задача присвоения меток класса (class label) наблюдениям (Observation) объектам из предметной области. Множество допустимых меток класса конечно. В свою очередь **класс** — это множество всех объектов с данным значением метки. Требуется построить алгоритм, способный классифицировать (присвоить метку) произвольный объект из исходного множества. Классификация, как правило, на этапе настройки использует обучение с учителем.



Байесовский классификатор

Широкий класс алгоритмов классификации, основанный на принципе максимума апостериорной вероятности. Для классифицируемого объекта вычисляются функции правдоподобия каждого из классов, по ним вычисляются апостериорные вероятности классов. Байесовский классификатор использует оценку апостериорного максимума (Maximum a posteriori estimation) для определения наиболее вероятного класса. Объект относится к тому классу, для которого апостериорная вероятность максимальна.

Байесовский подход к классификации основан на теореме, утверждающей, что если плотности распределения каждого из классов известны, то искомый алгоритм можно выписать в явном аналитическом виде. Более того, этот алгоритм оптимален, то есть обладает минимальной вероятностью ошибок.

На практике плотности распределения классов не известны. Их приходится оценивать (восстанавливать) по обучающей выборке. В результате байесовский алгоритм перестаёт быть оптимальным, так как восстановить плотность по выборке можно только с некоторой погрешностью. Чем короче выборка, тем выше шансы подогнать распределение под конкретные данные и столкнуться с эффектом переобучения.

К числу байесовских методов классификации относятся:

Наивный байесовский классификатор	Метод парзеновского окна
Линейный дискриминант Фишера	Метод радиальных базисных функций (RBF)
Квадратичный дискриминант	Логистическая регрессия

Модель наивного байесовского классификатора

Вероятностная модель для классификатора — это условная модель:

$$p(C|F_1, \dots, F_n)$$

над зависимой переменной класса *C* с малым количеством результатов или классов, зависимая от нескольких переменных $F_1, \dots, F_n$.

Используя теорему Байеса, запишем:

$$p(C \mid F_1, \dots, F_n) = \frac{p(C)\, p(F_1, \dots, F_n \mid C)}{p(F_1, \dots, F_n)}.$$

На практике интересен лишь числитель этой дроби, так как знаменатель не зависит от *C* и значения свойств F_i даны, так что знаменатель — константа.

Числитель эквивалентен совместной вероятности модели $p(C, F_1, \dots, F_n)$, которая может быть переписана, используя повторные приложения определений условной вероятности:

$$\begin{aligned} p(C, F_1, \dots, F_n) &= p(C)p(F_1, \dots, F_n|C) \\ &= p(C)p(F_1|C)p(F_2, \dots, F_n|C, F_1) \\ &= p(C)p(F_1|C)\, p(F_2|C, F_1)p(F_3, \dots, F_n|C, F_1, F_2) \\ &= p(C)p(F_1|C)\, p(F_2|C, F_1) \dots p(F_n|C, F_1, F_2, F_3, \dots, F_{n-1}) \end{aligned}$$

«Наивные» предположения условной независимости

Предположим, что каждое свойство F_i условно независимо от любого другого свойства F_j при $j \neq i$. Это означает, что если вероятность появления события $F_i|C$, не зависит от F_j, значит F_j можно отбросить

$$p(F_i|C, F_j) = p(F_i|C)$$

таким образом, совместная модель может быть выражена как:

$$p(C, F_1, \dots, F_n) = p(C) \cdot p(F_1|C) \cdot p(F_2|C) \cdot p(F_3|C) \cdot \dots \cdot p(F_n|C) = p(C) \prod_{i=1}^{n} p(F_i|C)$$

Это означает, что из предположения о независимости, условное распределение по классовой переменной *C* может быть выражено так:

$$p(C|F_1, \dots, F_n) = \frac{1}{Z} p(C) \prod_{i=1}^{n} p(F_i|C)$$

где $Z = p(F_1, \dots, F_n)$ — это масштабный множитель, зависящий только от $F_1, \dots, F_n$, то есть константа, если значения переменных известны.

Оценка параметров

Все параметры модели могут быть аппроксимированы относительными частотами из набора данных обучения. Это оценки максимального правдоподобия вероятностей. Непрерывные свойства, как правило, оцениваются через нормальное распределение. В качестве математического ожидания и дисперсии вычисляются статистики — среднее арифметическое и среднеквадратическое отклонение соответственно.

Если данный класс и значение свойства никогда не встречаются вместе в наборе обучения, тогда оценка, основанная на вероятностях, будет равна нулю. Это проблема, так как при перемножении нулевая оценка приведет к потере информации о других вероятностях. Поэтому предпочтительно проводить небольшие поправки во все оценки вероятностей так, чтобы никакая вероятность не была строго равна нулю.

Построение классификатора по вероятностной модели

Наивный байесовский классификатор объединяет модель с правилом решения. Одно общее правило должно выбрать наиболее вероятную гипотезу; оно известно *как апостериорное правило принятия решения* (maximum a posteriori - MAP). Соответствующий классификатор — это функция *classify*, определённая следующим образом:

$$Classify(f_1, \dots, f_n) = \arg\max_c \left(p(C = c) \cdot \prod_{i=1}^{n} p(F_i = f_i | C = c) \right)$$

Байесовская фильтрация спама

Байесовская фильтрация спама — метод для фильтрации спама, основанный на применении наивного байесовского классификатора, опирающегося на прямое использование теоремы Байеса.

История

Первой известной программой, фильтрующей почту с использованием байесовского классификатора, была программа iFile Джейсона Ренни, выпущенная в 1996 году. Программа использовала сортировку почты по папкам.

Первая академическая публикация по наивной байесовской фильтрации спама появилась в 1998 году. Вскоре после этой публикации была развернута работа по созданию коммерческих фильтров спама.

В 2002 г. Пол Грэм смог значительно уменьшить число ложноположительных срабатываний до такой степени, что байесовский фильтр мог использоваться в качестве единственного фильтра спама.

Описание

При обучении фильтра для каждого встреченного в письмах слова высчитывается и сохраняется его «вес» — оценка вероятности того, что письмо с этим словом — спам. В простейшем случае в качестве оценки используется частота: «появлений в спаме / появлений всего».

При проверке вновь пришедшего письма вероятность «спамовости» вычисляется по формуле ($Classify$) для множества гипотез.

$$P(B) = \sum_{i=1}^{N} P(A_i)P(B|A_i)$$

В данном случае «гипотезы» — это слова, и для каждого слова «достоверность гипотезы» $P(A_i) = N_{wordi}/N_{words\ total}$ — доля этого слова в письме, а «зависимость события от гипотезы» $P(B|A_i)$ — вычисленный ранее «вес» слова. То есть «вес» письма в данном случае — усреднённый «вес» всех его слов.

Отнесение письма к «спаму» или «не-спаму» производится по тому, превышает ли его «вес» некую планку, заданную пользователем (обычно берут 60-80 %). После принятия решения по письму в базе данных обновляются «веса» для вошедших в него слов.

Математические основы

Почтовые байесовские фильтры основываются на теореме Байеса. Теорема Байеса используется несколько раз в контексте спама:

- в первый раз, чтобы вычислить вероятность, что сообщение — спам, зная, что данное слово появляется в этом сообщении;
- во второй раз, чтобы вычислить вероятность, что сообщение — спам, учитывая все его слова (или соответствующие их подмножества);
- иногда в третий раз, когда встречаются сообщения с редкими словами.

Вычисление вероятности того, что сообщение, содержащее данное слово, является спамом:

$$\mathrm{P}(S|W) = \frac{\mathrm{P}(W|S) * \mathrm{P}(S)}{\mathrm{P}(W)} = \frac{\mathrm{P}(W|S) * \mathrm{P}(S)}{\mathrm{P}(W|S) * \mathrm{P}(S) + \mathrm{P}(W|H) * \mathrm{P}(H)}$$

$\mathrm{P}(S|W)$ — условная вероятность того, что сообщение—спам (S) , при условии, что слово *W*=«replica» находится в нём;

P (W) – полная вероятность того, что слово «Replica» содержится в сообщении

$\mathrm{P}(S)$ — полная вероятность того, что произвольное сообщение—спам;

$\mathrm{P}(W|S)$ — условная вероятность того, что слово «replica» появляется в сообщениях, если они являются спамом;

$\mathrm{P}(H)$ — полная вероятность того, что произвольное сообщение не спам H;

$\mathrm{P}(W|H)$ — условная вероятность того, что слово «replica» появляется в сообщениях, если они не спам (то есть «ham»).

Спамовость слова

Большинство байесовских программ обнаружения спама делают предположение об отсутствии априорных предпочтений у сообщения быть «spam» или «ham», и полагают, что у обоих случаев есть равные вероятности 50 %: P(*S*)=0.5, P(*H*)=0.5. О фильтрах, которые используют эту гипотезу, говорят как о фильтрах «без предубеждений» и данное предположение позволяет упрощать общую формулу до:

$$\mathrm{P}(S|W) = \frac{\mathrm{P}(W|S)}{\mathrm{P}(W|S) + \mathrm{P}(W|H)}$$

Значение $\mathrm{P}(S|W)$ называют «спамовостью» слова *W*;

$\mathrm{P}(W|S)$ приближённо равно относительной частоте сообщений, содержащих слово *W* и идентифицированных как спам во время фазы обучения:

$$\mathrm{P}(W_i|S) = \frac{count(M: W_i \epsilon M, M \epsilon S)}{\sum_j count(M: W_j \epsilon M, M \epsilon S)}$$

$\mathrm{P}(W|H)$ приближённо равно относительной частоте сообщений, содержащих слово *W* и идентифицированных как «ham» во время фазы обучения:

$$\mathrm{P}(W_i|H) = \frac{count(M: W_i \epsilon M, M \epsilon H)}{\sum_j count(M: W_j \epsilon M, M \epsilon H)}$$

Для того, чтобы эти приближения имели смысл, набор обучающих сообщений должен быть большим и достаточно представительным. Также желательно, чтобы набор обучающих сообщений соответствовал 50 % гипотезе о перераспределении между спамом и «ham», то есть что наборы сообщений «spam» и «ham» имели один и тот же размер.

Объединение индивидуальных вероятностей

Для решения задачи классификации сообщений лишь на 2 класса: *S* (спам) и *H* = ¬*S* («ham» - не спам) из теоремы Байеса можно вывести следующую формулу оценки вероятности «спамовости» всего сообщения, содержащего слова $W_1, W_2, \dots W_N$:

$$p(S|W_1, W_2, \dots W_N) = \text{ [по теореме Байеса]} = \frac{p(W_1,W_2,\dots W_N|S)*p(S)}{p(W_1,W_2,\dots W_N)} =$$

$$= \text{ [так как } W_i \text{ предполагаются независимыми]} = \frac{\prod_i p(W_i|S)*p(S)}{p(W_1,W_2,\dots W_N)} =$$

$$= \text{ [по теореме Байеса]} = \frac{\prod_i \frac{p(S|W_i)*p(W_i)}{p(S)} * p(S)}{p(W_1,W_2,\dots W_N)} =$$

$$= \text{ [по формуле полной вероятности]} = \frac{\prod_i \frac{p(S|W_i)*p(W_i)}{p(S|W_i)*p(W_i)} * p(S)}{\prod_i (p(W_i|S))*p(S) + \prod_i (p(W_i|\neg S))*p(\neg S)} =$$

$$= \frac{\prod_i (p(S|W_i)*p(W_i))*p(S)^{1-N}}{\prod_i (p(S|W_i)*p(W_i))*p(S)^{1-N} + \prod_i (p(\neg S|W_i)*p(W_i))*p(\neg S)^{1-N}} =$$

$$= \frac{\prod_i p(S|W_i)}{\prod_i (p(S|W_i)) + \left(\frac{p(\neg S)}{p(S)}\right)^{1-N} * \prod_i p(\neg S|W_i)}.$$

В результате

Таким образом, предполагая $p(S) = p(\neg S)$=0.5, имеем:

$$p = \frac{p_1 p_2 \dots p_N}{p_1 p_2 \dots p_N + (1 - p_1)(1 - p_2) \dots (1 - p_N)}$$

где:

- $p = p(S \mid W_1, W_2, \dots W_N)$ — вероятность, что сообщение, содержащее слова $W_1, W_2, \dots W_N$ — спам;
- p_1 — условная вероятность $p(S|W_1)$ того, что сообщение — спам, при условии, что оно содержит первое слово (к примеру, «replica»);
- p_2 — условная вероятность $p(S|W_2)$ того, что сообщение — спам, при условии, что оно содержит второе слово (к примеру, «watches»);
- p_N — условная вероятность $p(S|W_N)$ того, что сообщение — спам, при условии, что оно содержит N-е слово (к примеру, «home»).

Результат *p* обычно сравнивают с некоторым порогом (например, *0.5*, чтобы решить, является ли сообщение спамом или нет. Если p ниже, чем порог, сообщение рассматривают как вероятный «ham», иначе его рассматривают как вероятный спам.

Проблема редких слов

Она возникает в случае, если слово никогда не встречалось во время фазы обучения: и числитель, и знаменатель равны нулю, и в общей формуле, и в формуле спамовости.

В целом, слова, с которыми программа столкнулась только несколько раз во время фазы обучения, не являются репрезентативными (набор данных в выборке мал для того, чтобы сделать надёжный вывод о свойстве такого слова). Простое решение состоит в том, чтобы игнорировать такие ненадёжные слова.

«Нейтральные» слова — такие, как, «the», «a», «some», или «is» (в английском языке), или их эквиваленты на других языках — могут быть проигнорированы. Вообще говоря, некоторые байесовские фильтры просто игнорируют все слова, у которых спамовость около 0.5, так как в этом случае получается качественно лучшее решение. Учитываются только те слова, спамовость которых около 0.0 (отличительный признак законных сообщений — «ham»), или рядом с 1.0 (отличительный признаки спама).

Выводы о классификаторе спама

ПЛЮСЫ

- Прост
- Удобен
- Эффективен

МИНУСЫ

- Базируется на предположении, что одни слова чаще встречаются в спаме, а другие — в обычных письмах, и неэффективен, если данное предположение неверно
- Работает только с текстом

Спасибо за внимание