

Wykorzystanie uczenia maszynowego w AI

Uczenie się było ważną częścią AI od samego początku, ponieważ AI może naśladować ludzki poziom inteligencji. Osiągnięcie poziomu mimikry, który faktycznie przypomina uczenie się, zajęło dużo czasu i różnych podejść. Obecnie uczenie maszynowe może pochwalić się quasi-ludzkim poziomem uczenia się w określonych zadaniach, takich jak klasyfikacja obrazów lub przetwarzanie dźwięku, i stara się osiągnąć podobny poziom uczenia się w wielu innych zadaniach. Uczenie maszynowe nie jest całkowicie zautomatyzowane. Nie możesz powiedzieć komputerowi, żeby przeczytał książkę i oczekujesz, że cokolwiek zrozumie. Automatyzacja oznacza, że komputery mogą nauczyć się programować w celu wykonywania zadań zamiast czekać na ludzi, aby je zaprogramowali. Obecnie automatyzacja wymaga dużej ilości danych wybranych przez człowieka, a także analizy i szkolenia danych (ponownie pod nadzorem człowieka). To tak, jakby wziąć dziecko za rękę i pokierować jego pierwszym krokiem. Ponadto uczenie maszynowe ma inne ograniczenia, które są podyktowane sposobem uczenia się na podstawie danych. Każda rodzina algorytmów ma określone sposoby wykonywania zadań, a ten rozdział opisuje te metody. Celem jest zrozumienie, w jaki sposób AI podejmuje decyzje i prognozy. Podobnie jak w przypadku odkrywania człowieka za kurtyną w Czarnoksiężniku z Krainy Oz, w tym rozdziale odkrywasz maszynę i operatora stojącego za AI. Niemniej jednak nadal możesz cieszyć się niesamowitym uczuciem, widząc cudowne osiągnięcia, jakie może zapewnić uczenie maszynowe.

Podążanie wieloma różnymi drogami do nauki

Tak jak istoty ludzkie mają różne sposoby uczenia się od świata, tak naukowcy, którzy podeszli do problemu uczenia się AI, wybrali różne drogi. Każdy wierzył w konkretny przepis na inteligencję mimiczną. Do tej pory żaden pojedynczy model nie okazał się lepszy od innych. Twierdzenie o braku darmowego lunchu o konieczności zapłaty za każdą korzyść jest w pełni skuteczne. Każdy z tych wysiłków okazał się skuteczny w rozwiązywaniu konkretnych problemów. Ponieważ algorytmy są abstrakcyjne (patrz pasek boczny „Brak darmowego lunchu”), żaden algorytm nie jest lepszy od drugiego, chyba że udowodniono to w konkretnym, praktycznym problemie. Poniższe sekcje zawierają dodatkowe informacje na temat tej koncepcji używania różnych metod do nauki.

Odkrywanie pięciu głównych podejść do uczenia się AI

Algorytm jest rodzajem kontenera. Zapewnia pudełko do przechowywania metody rozwiązania określonego rodzaju problemu. Algorytmy przetwarzają dane przez szereg dobrze zdefiniowanych stanów. Stany nie muszą być deterministyczne, ale mimo to są zdefiniowane. Celem jest stworzenie wyniku, który rozwiąże problem. W niektórych przypadkach algorytm odbiera dane wejściowe, które pomagają zdefiniować dane wyjściowe, ale zawsze koncentruje się na danych wyjściowych. Algorytmy muszą wyrażać przejścia między stanami przy użyciu dobrze zdefiniowanego i formalnego języka zrozumiałego dla komputera. W przetwarzaniu danych i rozwiązywaniu problemu algorytm definiuje, poprawia i wykonuje funkcję. Funkcja jest zawsze specyficzna dla rodzaju problemu rozwiązywanego przez algorytm. Jak opisano w rozdziale „Unikanie sztucznej inteligencji” w Rozdziale 1, każde z pięciu plemion ma inną technikę i strategię rozwiązywania problemów, które skutkują unikalnymi algorytmami. Połączenie tych algorytmów powinno ostatecznie doprowadzić do nadrzędnego algorytmu, który będzie w stanie rozwiązać dany problem. Poniższe sekcje zawierają przegląd pięciu głównych technik algorytmicznych.

Rozumowanie symboliczne

Jedna z najwcześniejszych grup, symboliści, wierzyli, że wiedzę można uzyskać, operując symbolami (znakami oznaczającymi pewne znaczenie lub wydarzenie) i wywodząc od nich reguły. Łącząc złożone

systemy reguł, można uzyskać logiczną dedukcję wyniku, który chciałbyś poznać, w ten sposób symboliści ukształtowali swoje algorytmy, aby tworzyć reguły z danych. W symbolicznym rozumowaniu dedukcja poszerza sferę ludzkiej wiedzy, podczas gdy indukcja podnosi poziom wiedzy ludzkiej. Indukcja zwykle otwiera nowe pola eksploracji, podczas gdy dedukcja bada te pola.

Połączenia wzorowane na neuronach mózgu

Koneksjoniści są chyba najbardziej znanym z pięciu grup. Ta grupa stara się odtwarzać funkcje mózgu za pomocą krzemu zamiast neuronów. Zasadniczo każdy z neuronów (stworzony jako algorytm modelujący rzeczywisty odpowiednik) rozwiązuje niewielką część problemu, a równoległe użycie wielu neuronów rozwiązuje problem jako całość. Zastosowanie propagacji wstecznej lub wstecznej propagacji błędów ma na celu określenie warunków, w których błędy są usuwane z sieci zbudowanych tak, aby przypominały ludzkie neurony, poprzez zmianę wag (jak wiele danych wejściowych w wyniku) i stronniczości (które cechy są wybrano) sieci. Celem jest kontynuacja zmiany wag i odchyień, aż do momentu, gdy rzeczywista moc wyjściowa będzie zgodna z mocą docelową. W tym momencie sztuczny neuron strzela i przekazuje swój roztwór do następnego neuronu w linii.

Rozwiązanie stworzone przez tylko jeden neuron jest tylko częścią całego rozwiązania. Każdy neuron przekazuje informacje do następnego neuronu w linii, dopóki grupa neuronów nie wytworzy ostatecznego wyniku. Taka metoda okazała się najbardziej skuteczna w zadaniach podobnych do ludzkich, takich jak rozpoznawanie przedmiotów, rozumienie języka pisanego i mówionego oraz rozmawianie z ludźmi.

Algorytmy ewolucyjne testujące zmienność

Ewolucjoniści polegają na zasadach ewolucji w celu rozwiązywania problemów. Innymi słowy, strategia ta opiera się na przetrwaniu najsilniejszych (usunięcie wszelkich rozwiązań, które nie pasują do pożądanej wydajności). Funkcja sprawności określa żywotność każdej funkcji w rozwiązaniu problemu. Korzystając ze struktury drzewa, metoda rozwiązania szuka najlepszego rozwiązania na podstawie danych wyjściowych funkcji. Zwycięzca każdego poziomu ewolucji może zbudować funkcje następnego poziomu. Chodzi o to, że następny poziom zbliży się do rozwiązania problemu, ale może nie rozwiązać go całkowicie, co oznacza, że potrzebny jest kolejny poziom. Ta szczególna grupa w dużej mierze opiera się na rekursji i językach, które silnie wspierają rekursję w celu rozwiązania problemów. Ciekawym rezultatem tej strategii są ewoluujące algorytmy: jedna generacja algorytmów faktycznie tworzy następną generację.

Wnioskowanie bayesowskie

Grupa naukowców, zwana Bayesianami, dostrzegła, że niepewność była kluczowym aspektem, na który należy uważać, i że nauka nie była zapewniona, ale raczej miała miejsce jako ciągła aktualizacja poprzednich przekonań, które stawały się coraz bardziej dokładne. To spostrzeżenie skłoniło Bayesian do przyjęcia metod statystycznych, a zwłaszcza pochodnych z twierdzenia Bayesa, które pomagają obliczyć prawdopodobieństwa w określonych warunkach (na przykład, widząc kartę określonego ziarna, wartość początkową dla pseudolosowej sekwencji, losowane z talii po trzech innych kartach tego samego materiału siewnego).

Systemy uczące się przez analogię

Analogizatorzy wykorzystują maszyny jądra do rozpoznawania wzorców w danych. Rozpoznając wzorec jednego zestawu danych wejściowych i porównując go ze wzorcem znanego wyjścia, możesz stworzyć rozwiązanie problemu. Celem jest wykorzystanie podobieństwa w celu ustalenia najlepszego rozwiązania problemu. Jest to rodzaj rozumowania, który determinuje, że użycie konkretnego

rozwiązania działało w danych okolicznościach w pewnym momencie; dlatego stosowanie tego rozwiązania w podobnych okolicznościach powinno również działać. Jednym z najbardziej rozpoznawalnych wyników tego plemienia są systemy polecające. Na przykład, gdy kupujesz produkt na Amazon, system rekomendacji wymyśla inne powiązane produkty, które możesz chcieć kupić. Ostatecznym celem uczenia maszynowego jest połączenie technologii i strategii przyjętych przez pięć plemion w celu stworzenia jednego algorytmu (algorytmu nadrzędnego), który może nauczyć się wszystkiego. Oczywiście osiągnięcie tego celu jest daleko. Mimo to naukowcy tacy jak Pedro Domingos obecnie pracują do tego celu.

Zagłębianie się w trzy najbardziej obiecujące metody uczenia się AI

Późniejsze sekcje tego zawierają informacje na temat podstawowych algorytmów wybranych przez Bayesian, symbolistów i koneksjonistów. Grupy te reprezentują obecną i przyszłą granicę uczenia się na podstawie danych, ponieważ wywodzi się z nich jakkolwiek postęp w kierunku sztucznej inteligencji podobnej do człowieka, przynajmniej do czasu nowego przełomu z nowymi, bardziej niesamowitymi i potężnymi algorytmami uczenia się. Scenariusz uczenia maszynowego jest z pewnością znacznie większa niż te trzy algorytmy, ale w tym rozdziale skupiono się na tych trzech plemionach z powodu ich obecnej roli w sztucznej inteligencji. Oto streszczenie podejść :

* Naïve Bayes: Ten algorytm może być dokładniejszy niż lekarz w diagnozowaniu niektórych chorób. Ponadto ten sam algorytm może wykrywać spam i przewidywać nastroje na podstawie tekstu. Jest również szeroko stosowany w branży internetowej do łatwego przetwarzania dużych ilości danych.

* Sieci bayesowskie (forma wykresu): Ten wykres przedstawia złożoność świata pod względem prawdopodobieństwa.

* Drzewa decyzyjne: Algorytm drzewa decyzyjnego najlepiej reprezentuje symbolistów. Drzewo decyzyjne ma długą historię i wskazuje, w jaki sposób sztuczna inteligencja może podejmować decyzje, ponieważ przypomina serię zagnieżdżonych decyzji, które można narysować jako drzewo (stąd nazwa).

Te typy algorytmów są dalej podzielone na podkategorie. Na przykład drzewa decyzyjne są klasyfikowane jako drzewa regresji, drzewa klasyfikacyjne, drzewa wzmocnione, agregowane bootstrap i las rotacyjny. Możesz nawet zagłębić się w podtypy podkategorii. Losowy klasyfikator lasu jest rodzajem agregacji bootstrap, a stamtąd jest jeszcze więcej poziomów. Po przekroczeniu poziomów zaczniesz widzieć rzeczywiste algorytmy, które liczą się w tysiącach. Krótko mówiąc, ta książka zawiera przegląd nieskończenie bardziej złożonego tematu, który może wymagać wielu tomów do omówienia w najdrobniejszych szczegółach. Na wynos jest uchwycenie rodzaju algorytmu i nie popadanie w szczegóły.

W oczekiwaniu na kolejny przełom

W latach 80. XX wieku, kiedy systemy eksperckie rządziły sceną sztucznej inteligencji, większość naukowców i praktyków uważała uczenie maszynowe za niewielką gałąź sztucznej inteligencji, która koncentrowała się na uczeniu się, jak najlepiej odpowiadać na proste prognozy ze środowiska (reprezentowane przez dane) przy użyciu optymalizacji. Obecnie uczenie maszynowe ma przewagę nad sztuczną inteligencją, przewyższając systemy eksperckie w wielu aplikacjach i opracowaniach badawczych, oraz zasilając aplikacje sztucznej inteligencji, które naukowcy wcześniej uważali za niemożliwe przy takim poziomie dokładności i wydajności. Sieci neuronowe, rozwiązanie zaproponowane przez łączników, umożliwiły przełom w ciągu ostatnich kilku lat, wykorzystując połączenie zwiększonej pojemności sprzętowej, bardziej odpowiednich danych oraz wielu naukowców takich jak Geoffrey Hinton, Yann LeCun, Yoshua Bengio i wielu inni. Możliwości oferowane przez

algorytmy sieci neuronowej (nowe markowe głębokie uczenie się z powodu zwiększonej złożoności) rosną z dnia na dzień. Częste doniesienia prasowe opowiadają o nowych osiągnięciach w zakresie rozumienia audio, rozpoznawania obrazów i filmów, tłumaczenia języka, a nawet czytania warg. (Chociaż głębokie uczenie się nie ma wydajności HAL9000, zbliża się do ludzkiej wydajności) Ulepszenia są wynikiem intensywnego finansowania od dużych i małych firm w celu zaangażowania badaczy oraz dostępności potężnego oprogramowania, takiego jak Google TensorFlow i Microsoft Computational Network Toolkit, CNTK, które zapewnia zarówno naukowcom, jak i praktykom dostęp do technologii. Poszukaj jeszcze bardziej sensacyjnych innowacji AI w najbliższej przyszłości. Oczywiście badacze zawsze mogli ponownie uderzyć w ścianę, jak to miało miejsce w poprzednich zimach AI. Nikt nie może wiedzieć, czy AI osiągnie poziom ludzki przy użyciu obecnej technologii, czy ktoś odkryje główny algorytm, jak przewiduje Pedro Domingos, które rozwiążą wszystkie problemy związane z AI (niektóre z nich musimy jeszcze sobie wyobrazić). Niemniej uczenie maszynowe z pewnością nie jest modą napędzaną szumem; pozostanie tutaj, w obecnej, ulepszonej formie lub w postaci nowych algorytmów.

Poznawanie prawdy w prawdopodobieństwach

Niektóre strony internetowe uważają, że statystyki i uczenie maszynowe to dwie zupełnie różne technologie. Na przykład, gdy czytasz blog o nazwie Statystyka vs. uczenie maszynowe, walcz!, masz wrażenie, że te dwie technologie są nie tylko różne, ale wręcz wrogo nastawieni do siebie. Chociaż statystyki pokazują bardziej teoretyczne podejście do problemów, podczas gdy uczenie maszynowe opiera się wyłącznie na danych, statystyki i uczenie maszynowe mają ze sobą wiele wspólnego. Ponadto statystyki reprezentują jedną z pięciu grup (szkół myślenia), które umożliwiają uczenie maszynowe. Statystyki często wykorzystują prawdopodobieństwa - które są sposobem wyrażenia niepewności co do wydarzeń na świecie - podobnie jak uczenie maszynowe i sztuczna inteligencja (w większym stopniu niż czysta statystyka). Nie wszystkie problemy są jak gry w szachy lub Go, które pozwalają ci podjąć dużą, ale ograniczoną liczbę działań, kiedy zdecydujesz się je podjąć. Jeśli chcesz nauczyć się poruszać robotem w korytarzu wypełnionym ludźmi lub mieć samochód z własnym napędem z powodzeniem uczestniczącym w przejeździe, musisz wziąć pod uwagę, że plany (takie jak przejście z punktu A do punktu B) nie zawsze mają jeden wynik i że wiele wyników jest podatnych na osiadanie, każdy z innym prawdopodobieństwem. W pewnym sensie prawdopodobieństwo wspiera systemy AI w ich rozumowaniu, zapewniając wsparcie w podejmowaniu decyzji i dokonując, jak się wydaje, najlepszych, najbardziej racjonalnych wyborów pomimo niepewności. Niepewność może istnieć z różnych powodów, a SI należy uświadomić poziom niepewności poprzez skuteczne wykorzystanie prawdopodobieństwa:

1. Niektóre sytuacje nie są pewne, ponieważ mają charakter losowy. Podobne sytuacje są z natury stochastyczne. Na przykład w grach karcianych nie możesz mieć pewności, jaką kartę będziesz mieć po tym, jak krupier tasuje i rozdaje karty.
2. Nawet jeśli sytuacja nie jest przypadkowa, nieprzebranie wszystkich jej aspektów (niekompletna obserwacja) powoduje niepewność co do tego, jak wszystko się potoczy. Na przykład robot idący korytarzem zatłoczonym ludźmi nie może znać zamierzonego kierunku każdej osoby (nie może czytać w ich myślach), ale może sformułować przypuszczenie na podstawie częściowej obserwacji ich zachowania. Jak przy każdym zgadywaniu, robot ma szansę mieć rację i się mylić.
3. Ograniczenia w sprzęcie rejestrującym dane światowe (zwane czujnikami) i przybliżenia w przetwarzaniu danych mogą sprawić, że wyniki uzyskane z takich danych będą niepewne. Pomiar często są obciążone błędami ze względu na stosowane narzędzia i sposób wykonywania pomiaru. Ponadto ludzie są często narażeni na uprzedzenia poznawcze i łatwo padają ofiarą złudzeń lub

martwych punktów. Podobnie AI jest ograniczona jakością odbieranych danych. Przybliżenia i błędy wprowadzają niepewność do każdego algorytmu.

Określanie co prawdopodobieństwo może zrobić

Prawdopodobieństwo mówi ci o prawdopodobieństwie zdarzenia i wyrażasz je jako liczbę. Na przykład, jeśli rzucisz monetę w powietrze, nie wiesz, czy wyląduje ona jako orzeł czy reszka, ale możesz określić prawdopodobieństwo obu wyników. Prawdopodobieństwo zdarzenia mierzy się w zakresie od 0 (brak prawdopodobieństwa wystąpienia zdarzenia) do 1 (pewność wystąpienia zdarzenia). Wartości pośrednie, takie jak 0,25, 0,5 i 0,75, mówią, że zdarzenie nastąpi z pewną częstotliwością, gdy spróbuje się wystarczająco dużo razy. Jeśli pomnożysz prawdopodobieństwo przez liczbę całkowitą reprezentującą liczbę prób, które zamierzasz spróbować, otrzymasz oszacowanie, ile razy zdarzenie powinno zdarzyć się średnio, jeśli wszystkie próby zostaną wypróbowane. Na przykład, jeśli zdarzenie wystąpi z prawdopodobieństwem $p = 0,25$ i spróbujesz 100 razy, najprawdopodobniej będziesz świadkiem, że zdarzenie to nastąpi $0,25 * 100 = 25$ razy. Tak się składa, że wynikiem $p = 0,25$ jest prawdopodobieństwo wybrania określonego koloru przy losowym wybieraniu karty z talii kart. Francuskie karty do gry stanowią klasyczny przykład wyjaśniania prawdopodobieństw. Talia zawiera 52 karty równo podzielone na cztery kolory: trefl i piki, które są czarne, oraz karo i kier, które są czerwone. Jeśli więc chcesz ustalić prawdopodobieństwo wybrania asa, musisz wziąć pod uwagę, że istnieją cztery asy różnych kolorów. Odpowiedź pod względem prawdopodobieństwa wynosi $p = 4/52 = 0,077$. Prawdopodobieństwa wynoszą od 0 do 1; żadne prawdopodobieństwo nie może przekroczyć takich granic. Prawdopodobieństwa definiujesz empirycznie na podstawie obserwacji. Po prostu policz, ile razy wydarzenie ma miejsce w odniesieniu do wszystkich wydarzeń, które Cię interesują. Załóżmy na przykład, że chcesz obliczyć prawdopodobieństwo, ile razy zdarzy się oszustwo podczas transakcji bankowych lub ile razy ludzie zapadną na określoną chorobę w danym kraju. Po obejrzeniu zdarzenia możesz oszacować prawdopodobieństwo z nim związane, zliczając liczbę jego wystąpień i dzieląc przez całkowitą liczbę zdarzeń.

Możesz policzyć liczbę przypadków oszustwa lub choroby, używając zarejestrowanych danych (głównie pobranych z baz danych), a następnie podzielić tę liczbę przez całkowitą liczbę dostępnych ogólnych zdarzeń lub obserwacji. Dlatego dzielisz liczbę oszustw przez liczbę transakcji w ciągu roku lub liczysz liczbę osób, które zachorowały w ciągu roku w stosunku do liczby ludności określonego obszaru. Wynikiem jest liczba z zakresu od 0 do 1, którą można wykorzystać jako podstawowe prawdopodobieństwo dla określonego zdarzenia w określonych okolicznościach.

Zliczanie wszystkich wystąpień zdarzenia nie zawsze jest możliwe, dlatego musisz wiedzieć o próbkowaniu. Próbkowanie, które jest czynnością opartą na pewnych oczekiwaniach prawdopodobieństwa, pozwala zaobserwować niewielką część większego zestawu zdarzeń lub obiektów, a jednocześnie być w stanie określić prawidłowe prawdopodobieństwa zdarzenia, a także dokładne miary, takie jak pomiary ilościowe lub jakościowe klasy związane z zestawem obiektów. Na przykład jeśli chcesz śledzić sprzedaż samochodów w Stanach Zjednoczonych w ostatnim miesiącu, nie musisz śledzić każdej sprzedaży w tym kraju. Korzystając z próbki obejmującej sprzedaż od kilku sprzedawców samochodów w całym kraju, możesz określić miary ilościowe, takie jak średnia cena sprzedanego samochodu lub miary jakościowe, takie jak najczęściej sprzedawany model samochodu.

Biorąc pod uwagę wcześniejszą wiedzę

Prawdopodobieństwo ma sens pod względem czasu i przestrzeni, ale niektóre inne warunki również wpływają na prawdopodobieństwo pomiaru. Kontekst jest ważny. Gdy oszacujesz prawdopodobieństwo zdarzenia, możesz (czasem błędnie) wierzyć, że możesz zastosować obliczone prawdopodobieństwo do każdej możliwej sytuacji. Terminem wyrażającym to przekonanie jest

prawdopodobieństwo a priori, co oznacza ogólne prawdopodobieństwo zdarzenia. Na przykład, gdy rzucisz monetą, jeśli moneta jest uczciwa, prawdopodobieństwo a priori orła wynosi około 50 procent (gdy zakładasz również istnienie niewielkiego prawdopodobieństwa wylądowania monety na jej krawędzi). Bez względu na to, ile razy rzucisz monetą, w przypadku nowego rzutu prawdopodobieństwo dla głów wynosi nadal około 50 procent. Jednak w niektórych innych sytuacjach, jeśli zmienisz kontekst, prawdopodobieństwo a priori nie jest już ważne, ponieważ wydarzyło się coś subtelnego i to zmieniło. W takim przypadku możesz wyrazić to przekonanie jako prawdopodobieństwo a posteriori, które jest prawdopodobieństwem a priori po tym, jak coś się zmieniło. Na przykład prawdopodobieństwo, że dana osoba jest kobietą, wynosi z grubsza około 50 procent. Jednak prawdopodobieństwo może się znacznie różnić, jeśli weźmie się pod uwagę tylko określone przedziały wiekowe, ponieważ kobiety zwykle żyją dłużej, a po pewnym wieku starsza grupa wiekowa zawiera więcej kobiet niż mężczyźn. Jako kolejny przykład związany z płcią, kobiety na ogół przewyższają mężczyzn na głównych uniwersytetach. Dlatego, biorąc pod uwagę te dwa konteksty, prawdopodobieństwo a posteriori jest inne niż oczekiwane z góry. Pod względem rozkładu płci natura i kultura mogą stwarzać inne prawdopodobieństwo a posteriori. Poniższe sekcje pomagają bardziej szczegółowo zrozumieć przydatność prawdopodobieństwa.

Prawdopodobieństwo warunkowe i Naïve Bayes

Możesz wyświetlić przypadki, takie jak związane z płcią, wymienione w poprzedniej sekcji jako prawdopodobieństwo warunkowe i wyrażaj to jako $p(y | x)$, które odczytujesz jako prawdopodobieństwo wystąpienia zdarzenia y , biorąc pod uwagę, że wystąpiło x . Prawdopodobieństwa warunkowe są bardzo potężnym narzędziem do uczenia maszynowego i sztucznej inteligencji. W rzeczywistości, jeśli prawdopodobieństwo a priori może się tak bardzo zmienić z powodu pewnych okoliczności, znajomość możliwych okoliczności może zwiększyć twoje szanse na prawidłowe przewidywanie zdarzenia, obserwując przykłady - co jest dokładnie tym, co ma robić uczenie maszynowe. Na przykład, jak wspomniano wcześniej, oczekiwanie, że przypadkowa osoba będzie mężczyzną lub kobietą, wynosi na ogół 50 procent. Ale co, jeśli dodasz dowód, że włosy tej osoby są długie lub krótkie? Możesz oszacować prawdopodobieństwo posiadania długich włosów na 35 procent populacji; jednak jeśli obserwujesz tylko populację kobiet, prawdopodobieństwo wzrasta do 60 procent. Jeśli odsetek jest tak wysoki w populacji kobiet, w przeciwieństwie do prawdopodobieństwa a priori, algorytm uczenia maszynowego, zwany Naïve Bayes, wymaga danych wejściowych wskazujących, czy włosy danej osoby są długie, czy krótkie. W rzeczywistości algorytm Naïve Bayes korzysta ze zwiększenia szansy na poprawną prognozę, znając okoliczności otaczające prognozę. Wszystko zaczyna się od wielobnego Bayesa i jego rewolucyjnego twierdzenia o prawdopodobieństwach. W rzeczywistości, jak zauważono gdzie indziej, w książce jedno z plemion uczących się maszyn nosi imię jego (Bayesianie). Bayesianie stosują różne metody statystyczne do rozwiązywania problemów, wszystkie oparte na obserwacji prawdopodobieństwa pożądanego wyniku we właściwym kontekście, przed i po zaobserwowaniu samego wyniku.

W oparciu o te obserwacje rozwiązują problem wschodu słońca (szacując prawdopodobieństwo, że Słońce wstanie jutro), łącząc kolejne obserwacje i stale aktualizując swoje oszacowanie prawdopodobieństwa wschodu Słońca proporcjonalnie do liczby razy, gdy byli świadkami długich serii o świtach wcześniej. Badacze danych mają duże oczekiwania w zakresie opracowania zaawansowanych algorytmów opartych na prawdopodobieństwie bayesowskim. Magazyn MIT Technology Review wspomina o uczeniu maszynowym Bayesian jako o nowej technologii, która zmieni nasz świat. Jednak twierdzenie Bayesa o fundamentach nie jest wcale takie skomplikowane (choć może być trochę sprzeczne z intuicją, jeśli normalnie bierze się pod uwagę, jak większość ludzi, tylko prawdopodobieństwa a priori bez uwzględnienia prawdopodobieństwa a posteriori).

Biorąc pod uwagę twierdzenie Bayesa

Oprócz bycia prezbiteriańskim pastorem, wielbny Thomas Bayes był także statystykiem i filozofem, który sformułował swoje twierdzenie w pierwszej połowie XVIII wieku. Twierdzenie nigdy nie zostało opublikowane, gdy żył. Jego publikacja zrewolucjonizowała teorię prawdopodobieństwa, wprowadzając ideę prawdopodobieństwa warunkowego wspomnianą w poprzednim rozdziale. Dzięki twierdzeniu Bayesa przewidywanie prawdopodobieństwa bycia mężczyzną lub kobietą staje się łatwiejsze, jeśli dowodem jest to, że dana osoba ma długie włosy. Oto formuła stosowana przez Thomasa Bayesa:

$$P(B | E) = P(E | B) * P(B) / P(E)$$

Wielbny Bayes nie wymyślił Naïve Bayes; sformułował jedynie twierdzenie. W rzeczywistości nie ma pewności co do przypisania algorytmu. Po raz pierwszy pojawił się w podręczniku w 1973 r. Bez odniesienia do jego twórcy i przeszedł niezauważony przez ponad dekadę, dopóki w 1990 r. naukowcy nie zauważyli, jak wykonuje niewiarygodnie dokładne prognozy, jeśli jest zasilany wystarczająco dokładnymi danymi. Czytanie formuły przy użyciu poprzedniego przykładu jako danych wejściowych może zapewnić lepsze zrozumienie formuły sprzecznej z intuicją:

$P(B | E)$: Prawdopodobieństwo przekonania (B) na podstawie zestawu dowodów (E) (prawdopodobieństwo wsteczne). Czytaj wiarę jako alternatywny sposób wyrażenia hipotezy. W tym przypadku hipoteza jest taka, że osoba jest kobietą, a dowodem są długie włosy. Znając prawdopodobieństwo takiego przekonania, dane dowody mogą pomóc w przewidywaniu płci danej osoby e.

$P(E | B)$: Prawdopodobieństwo posiadania długich włosów, gdy osoba jest płci żeńskiej. Termin ten odnosi się do prawdopodobieństwa dowodów w podgrupie, co samo w sobie jest prawdopodobieństwem warunkowym. W tym przypadku liczba wynosi 60 procent, co przekłada się na wartość 0,6 we wzorze (wcześniejsze prawdopodobieństwo).

$P(B)$: Ogólne prawdopodobieństwo bycia kobietą; to znaczy prawdopodobieństwo przekonania a priori. W takim przypadku prawdopodobieństwo wynosi 50 procent lub wartość 0,5 (prawdopodobieństwo).

$P(E)$: Ogólne prawdopodobieństwo posiadania długich włosów. Oto kolejne prawdopodobieństwo a priori, tym razem związane z zaobserwowanymi dowodami. W tym wzorze prawdopodobieństwo wynosi 35 procent, co stanowi wartość 0,35 (dowód).

Jeśli rozwiążesz poprzedni problem za pomocą formuły Bayesa i wartości, które wyróżniłeś, wynikiem będzie 06. * 0,5 / 0,35 = 0,857. Jest to wysoki odsetek prawdopodobieństwa, co prowadzi do potwierdzenia, że biorąc pod uwagę takie dowody, osoba ta jest prawdopodobnie kobietą. Innym częstym przykładem, który może podnieść brwi i który jest rutynowo spotykany w podręcznikach i czasopismach naukowych, jest pozytywny test medyczny. Jest to dość interesujące dla lepszego zrozumienia, w jaki sposób prawdopodobieństwa wcześniejsze i późniejsze mogą się bardzo zmienić w różnych okolicznościach. Powiedz, że martwisz się, że masz rzadką chorobę występującą u 1% populacji. Zdajesz test, a wyniki są pozytywne. Testy medyczne nigdy nie są idealnie dokładne, a laboratorium mówi ci, że gdy jesteś chory, test jest pozytywny w 99 procentach przypadków, podczas gdy jesteś zdrowy, test będzie negatywny w 99 procentach przypadków. Teraz, korzystając z tych danych, od razu uważasz, że jesteś chory, biorąc pod uwagę wysoki odsetek pozytywnych testów, gdy dana osoba jest chora (99 procent). Jednak rzeczywistość jest zupełnie inna. W tym przypadku liczby, które należy podłączyć do twierdzenia Bayesa, są następujące:

0,99 jako P (E | B)

0,01 jako P (B)

$0,01 * 0,99 + 0,99 * 0,01 = 0,0198$ jako P (E)

Obliczenia wynoszą wtedy $0,01 * 0,99 / 0,0198 = 0,5$, co odpowiada jedynie 50-procentowemu prawdopodobieństwu zachorowania. W końcu twoje szanse na zachorowanie są większe niż się spodziewałeś. Możesz się zastanawiać, jak to możliwe. Faktem jest, że liczba osób widzących pozytywną odpowiedź z testu jest następująca:

Kto jest chory i otrzymuje prawidłową odpowiedź z testu: Ta grupa jest prawdziwie pozytywna i stanowi 99 procent z 1 procent

populacja, która zachoruje. Kto nie jest chory i dostaje złą odpowiedź z testu: ta grupa to 1 procent z 99 procent populacji, która otrzymuje pozytywną odpowiedź, mimo że nie jest chora. Ponownie jest to mnożenie 99 procent i 1 procent. Ta grupa odpowiada wynikom fałszywie dodatnim. Jeśli spojrzysz na problem z tej perspektywy, staje się jasne, dlaczego. Ograniczając kontekst do osób, które otrzymają pozytywną odpowiedź na test, prawdopodobieństwo bycia w grupie prawdziwych pozytywnych jest takie samo jak prawdopodobieństwo bycia w fałszywych pozytywnych.

Wyobrażanie świata jako wykresu

Twierdzenie Bayesa może pomóc ci wydedukować, jak prawdopodobne jest, że coś się wydarzy w określonym kontekście, w oparciu o ogólne prawdopodobieństwa samego faktu i dowodów, które badasz, w połączeniu z prawdopodobieństwem dowodów, biorąc pod uwagę ten fakt. Rzadko pojedynczy materiał dowodowy zmniejsza wątpliwości i zapewni wystarczającą pewność prognozy, aby zapewnić, że tak się stanie. Jako prawdziwy detektyw, aby osiągnąć pewność, musisz zebrać więcej dowodów i sprawić, by poszczególne elementy współpracowały podczas dochodzenia. Zauważenie, że dana osoba ma długie włosy, nie wystarczy, aby ustalić, czy dana osoba jest kobietą czy mężczyzną. Dodanie danych o wzroście i wadze może zwiększyć pewność siebie. Algorytm Naïve Bayes pomaga uporządkować wszystkie zebrane dowody i uzyskać bardziej solidne prognozy z większym prawdopodobieństwem bycia poprawnym. Zebrane dowody rozpatrywane osobno nie mogą uchronić Cię przed ryzykiem niepoprawnego przewidywania, ale wszystkie zebrane dowody mogą osiągnąć bardziej ostateczne rozwiązanie. Poniższy przykład pokazuje, jak działają rzeczy w klasyfikacji Naïve Bayes. Jest to stary, znany problem, ale reprezentuje on możliwości, których można oczekiwać od AI. Zbiór danych pochodzi z artykułu „Indukcja drzew decyzyjnych” Johna Rossa Quinlana. Quinlan jest informatykiem, który przyczynił się do rozwoju innego algorytmu uczenia maszynowego, drzewa decyzyjne w fundamentalny sposób, ale jego przykład działa dobrze z każdym rodzajem algorytmu uczenia się. Problem wymaga od AI odgadnięcia najlepszych warunków do gry w tenisa, biorąc pod uwagę warunki pogodowe. Zestaw funkcji opisanych przez Quinlan jest następujący:

Perspektywy: słonecznie, pochmurno lub deszczowo

Temperatura: chłodna, łagodna lub gorąca

Wilgotność: wysoka lub normalna

Wietrznie: prawda lub fałsz

Poniższa tabela zawiera pozycje bazy danych użyte w przykładzie:

Widok: Temperatura : Wilgotność : Wietrznie : Gra

Słonecznie : Gorąco : Wysoka : Fałsz : Nie

Słonecznie : Gorąco : Wysoka : Prawda : Nie

Pochmurno : Gorąco : Wysoka : Fałsz : Tak

Deszczowo : Łagodna : Wysoka : Fałsz : Tak

Deszczowo : Chłodno : Normalna : Fałsz : Tak

Deszczowo : Chłodno : Normalna : Prawda : Nie

Pochmurno : Chłodno Normalna : Prawda : Tak

Słonecznie : Łagodna : Wysoka : Fałsz : Nie

Słonecznie : Chłodno : Normalna : Fałsz : Tak

Deszczowo : Łagodna : Normalna : Fałsz : Tak

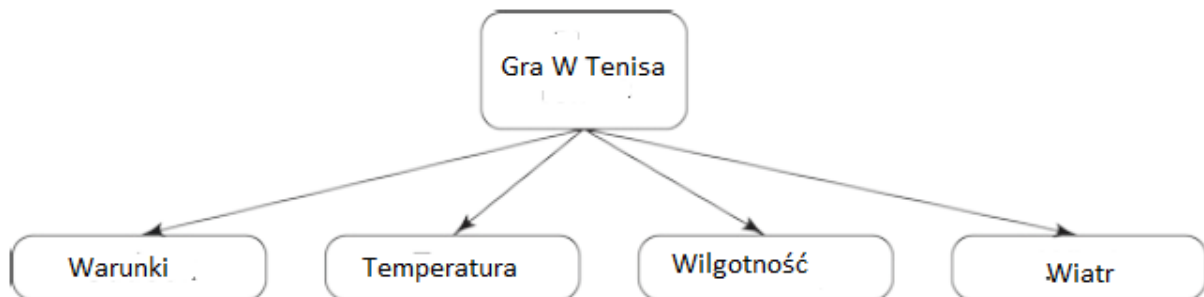
Słonecznie : Łagodna : Normalna : Prawda : Tak

Pochmurno : Łagodna : Wysoka : Prawda : Tak

Pochmurno : Gorąco : Normalna : Fałsz : Tak

Deszczowo : Łagodna : Wysoka : Prawda : Nie

Możliwość gry w tenisa zależy od czterech argumentów pokazanych na rysunku



Rezultatem tego przykładu nauki AI jest decyzja, czy grać w tenisa, biorąc pod uwagę warunki pogodowe (dowody). Używanie samego widoku (słonecznie, pochmurno lub deszczowo) nie wystarczy, ponieważ temperatura i wilgotność mogą być zbyt wysokie lub wiatr może być silny. Argumenty te reprezentują rzeczywiste warunki, które mają wiele przyczyn lub przyczyny, które są ze sobą powiązane. Algorytm Naïve Bayes ma umiejętność prawidłowego zgadywania, gdy istnieje wiele przyczyn. Algorytm oblicza wynik na podstawie prawdopodobieństwa podjęcia konkretnej decyzji i pomnożony przez prawdopodobieństwo dowodów związanych z tą decyzją. Na przykład, aby ustalić, czy grać w tenisa, gdy pogoda jest słoneczna, ale wiatr jest silny, algorytm oblicza wynik dla pozytywnej odpowiedzi, mnożąc ogólne prawdopodobieństwo gry (9 rozegranych gier z 14 występów) przez prawdopodobieństwo dzień jest słoneczny (2 z 9 rozegranych gier) i wietrzne warunki podczas gry w tenisa (3 z 9 rozegranych gier). Te same zasady obowiązują w przypadku negatywnym (który ma różne prawdopodobieństwa, że nie będzie grał w określonych warunkach):

prawdopodobieństwo gry: $9/14 * 2/9 * 3/9 = 0,05$

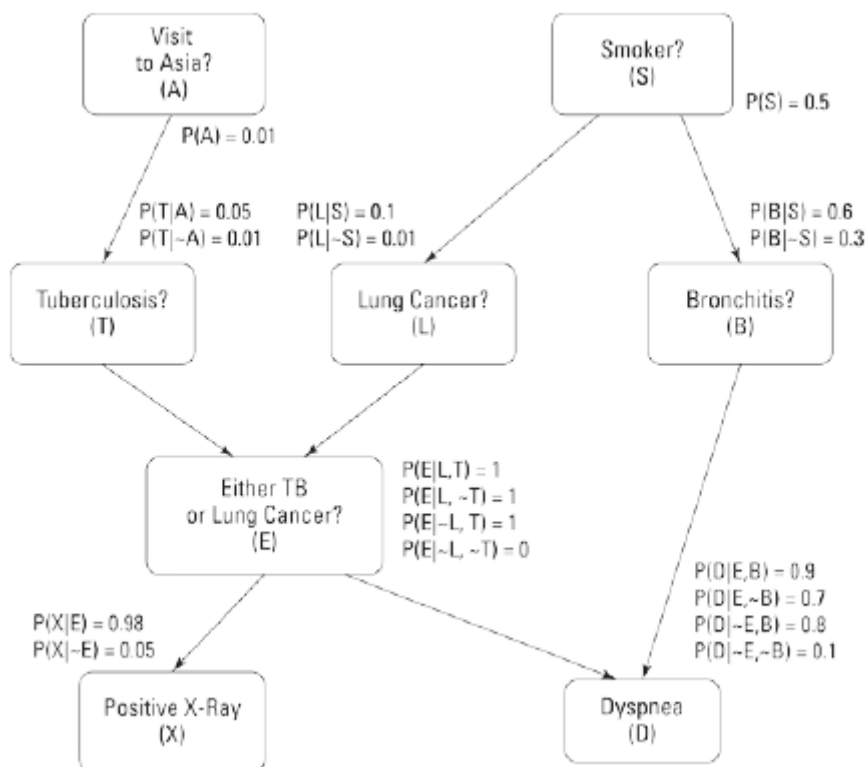
prawdopodobieństwo braku gry: $5/14 * 3/5 * 3/5 = 0,13$

Ponieważ wynik prawdopodobieństwa jest wyższy, algorytm decyduje, że bezpieczniej jest nie grać w takich warunkach. Oblicza takie prawdopodobieństwo poprzez zsumowanie dwóch wyników i podzielenie obu wyników przez ich sumę:

prawdopodobieństwo gry: $0,05 / (0,05 + 0,13) = 0,278$

prawdopodobieństwo braku gry: $0,13 / (0,05 + 0,13) = 0,722$

Możesz dodatkowo rozszerzyć Naïve Bayes, aby reprezentował relacje bardziej złożone niż szereg czynników wskazujących na prawdopodobieństwo wyniku za pomocą sieci bayesowskiej, która składa się z wykresów pokazujących, jak zdarzenia wpływają na siebie nawzajem. Wykresy bayesowskie mają węzły reprezentujące zdarzenia i łuki pokazujące, które zdarzenia wpływają na inne, wraz z tabelą prawdopodobieństw warunkowych, które pokazują, jak relacja działa pod względem prawdopodobieństwa. Rysunek pokazuje słynny przykład sieci bayesowskiej zaczerpniętej z pracy naukowej z 1988 r. „Lokalne obliczenia z prawdopodobieństwami na strukturach graficznych i ich zastosowaniu w systemach eksperckich” autorstwa Lauritzena, Steffena L. i Davida J. Spiegelhaltera, opublikowanych przez Journal of the Royal Statistics Society



Przedstawiona sieć nazywa się Asia. Pokazuje możliwe stany pacjenta i co go powoduje. Na przykład, jeśli pacjent ma duszność, może to być skutek gruźlicy, raka płuc lub zapalenia oskrzeli. Wiedza, czy pacjent pali, przebywa w Azji, czy ma nieprawidłowe wyniki badań rentgenowskich (daje pewność pewnych dowodów, a priori w języku bayesowskim) pomaga ustalić rzeczywiste (późniejsze) prawdopodobieństwo wystąpienia którejkolwiek z patologii w wykres. Sieci bayesowskie, choć intuicyjne, mają za sobą złożoną matematykę i są potężniejsze niż prosty algorytm Naïve Bayesa,

ponieważ naśladują świat jako sekwencję przyczyn i skutków opartych na prawdopodobieństwie. Sieci bayesowskie są tak skuteczne, że można je wykorzystać do przedstawienia każdej sytuacji. Mają różne zastosowania, takie jak diagnozy medyczne, łączenie niepewnych danych pochodzących z wielu czujników, modelowanie ekonomiczne i monitorowanie złożonych systemów, takich jak samochód. Na przykład, ponieważ jazda w ruchu drogowym może wiązać się ze złożonymi sytuacjami z wieloma pojazdami, konsorcjum Analysis of Massive Data Streams (AMIDST), we współpracy z producentem samochodów Daimler, opracowało sieć Bayesian, która może rozpoznawać manewry innych pojazdów i zwiększać bezpieczeństwo jazdy.

Rosnące drzewa, które można sklasyfikować

Drzewo decyzyjne to inny rodzaj kluczowego algorytmu uczenia maszynowego, który wpływa na implementację sztucznej inteligencji i uczenie się. Algorytmy drzewa decyzyjnego nie są nowe, ale mają długą historię. Pierwszy tego rodzaju algorytm pochodzi z lat 70. XX wieku (z wieloma późniejszymi wariantami). Gdy weźmie się pod uwagę eksperymenty i oryginalne badania, użycie drzew decyzyjnych sięga jeszcze wcześniej - nawet wstecz. Jako główny algorytm symboliczny drzewa decyzyjne cieszą się dużą popularnością, ponieważ są intuicyjnym rodzajem algorytmu. Łatwo jest przełożyć dane wyjściowe na reguły, a tym samym ułatwić zrozumienie wyników przez ludzi. Drzewa decyzyjne są również niezwykle łatwe w użyciu. Wszystkie te cechy sprawiają, że są one skutecznym i atrakcyjnym bez zastanowienia w stosunku do modeli, które wymagają złożonych transformacji macierzy danych wejściowych lub niezwykle dokładnego strojenia hiperparametrów. Symbolizm to podejście oparte na sztucznej inteligencji oparte na logicznych stwierdzeniach i szerokim zastosowaniu dedukcji. Odliczenie poszerza wiedzę z czego wiemy i indukcja formułuje ogólne zasady, zaczynając od dowodów.

Prognozowanie wyników przez podział danych

Jeśli masz grupę miar i chcesz opisać je za pomocą pojedynczej liczby, użyj średniej arytmetycznej (sumując wszystkie miary i dzieląc przez liczbę miar). W podobny sposób, jeśli masz grupę klas lub cech (na przykład masz zestaw danych zawierający zapisy wielu ras psów lub rodzajów produktów), możesz użyć najczęstszej klasy w grupie, aby przedstawić je wszystkie, który nazywa się trybem. Tryb jest kolejną miarą statystyczną, taką jak średnia, ale zawiera wartość (miarę lub klasę), która pojawia się najczęściej. Zarówno średnia, jak i tryb starają się zgłosić liczbę lub klasę, która zapewnia największą pewność odgadnięcia następnego elementu grupy, ponieważ powodują one najmniej błędów. W pewnym sensie są to predyktory, które uczą się odpowiedzi na podstawie istniejących danych. Drzewa decyzyjne wykorzystują środki i tryby jako predyktory, dzieląc zestaw danych na mniejsze zbiory, których środki lub tryby są najlepszymi możliwymi predyktorami dla danego problemu.

Podział problemu w celu łatwego znalezienia rozwiązania jest również powszechną strategią w wielu algorytmach dzielenia i zwyciężaj. Podobnie jak w przypadku armii wroga w bitwie, jeśli możesz rozdzielić wroga i walczyć z nim pojedynczo, możesz osiągnąć łatwiejsze zwycięstwo. Wykorzystując próbkę obserwacji jako punkt wyjścia, algorytm wyszukuje reguły, które generowały klasy wyjściowe (lub wartości liczbowe podczas pracy z problemem regresji), dzieląc macierz wejściową na coraz mniejsze partycje, dopóki proces nie uruchomi reguły zatrzymywania. Takie cofanie się od reguł ogólnych do ogólnych jest typowe dla ludzkiej dedukcji, traktowanej przez logikę i filozofię. W kontekście uczenia maszynowego takie odwrotne rozumowanie uzyskuje się poprzez przeszukanie wszystkich możliwych sposobów podzielenia szkolenia w próbie i zdecydowanie, w zachłanny sposób, zastosować podział, który maksymalizuje pomiary statystyczne na wynikowych partycjach. Algorytm jest chciwy, gdy zawsze dokonuje wyboru, aby zmaksymalizować wynik w bieżącym etapie procesu optymalizacji, niezależnie od tego, co może się wydarzyć w kolejnych krokach. W rezultacie chciwy algorytm może nie osiągnąć globalnej optymalizacji. Podział występuje w celu egzekwowania prostej zasady: każda

partycja danych początkowych musi ułatwiać przewidywanie wyniku docelowego, który charakteryzuje się innym i bardziej korzystnym rozkładem klas (lub wartości) niż pierwotna próbka. Algorytm tworzy partycje, dzieląc dane. Określa podział danych, najpierw oceniając funkcje. Następnie ocenia wartości funkcji, które mogą przynieść maksymalne ulepszenie specjalnej miary statystycznej - to znaczy miary, która odgrywa rolę funkcji kosztu w drzewie decyzyjnym. Wiele pomiarów statystycznych określa, jak wykonać podziały w drzewie decyzyjnym. Wszyscy trzymają się idei, że podział musi poprawić się na oryginalnej próbce lub innym możliwym podziale, jeśli dzięki temu przewidywanie jest bezpieczniejsze. Do najczęściej używanych pomiarów należą zanieczyszczenie gini, przyrost informacji i redukcja wariancji (w przypadku problemów z regresją). Pomiar te działają podobnie, dlatego ta część koncentruje się na pozyskiwaniu informacji, ponieważ jest to najbardziej intuicyjny pomiar i pokazuje, w jaki sposób drzewo decyzyjne może wykryć zwiększoną zdolność predykcyjną (lub zmniejszone ryzyko) w najprostszy sposób dla pewnego podziału. Ross Quinlan stworzył algorytm drzewa decyzyjnego na podstawie zdobywanie informacji (ID3) w latach siedemdziesiątych XX wieku i wciąż jest dość popularny dzięki niedawno zaktualizowanej wersji do C4.5. Zdobywanie informacji zależy od wzoru na entropię informacyjną (opracowany przez Claude'a Shannona, amerykańskiego matematyka i inżyniera znanego jako ojciec teorii informacji), uogólnione sformułowanie, które opisuje oczekiwaną wartość z informacji zawartych w komunikacie:

Entropia Shannona $E = -\sum (p(i) \times \log_2(p(i)))$

W formule rozważasz wszystkie klasy pojedynczo i sumujesz wynik mnożenia każdej z nich. W mnożeniu, które musi przyjąć każda klasa, $p(i)$ jest prawdopodobieństwem dla tej klasy (wyrażonym w zakresie od 0 do 1), a \log_2 jest logarytmem podstawowym 2. Zaczynając od próbki, w której chcesz sklasyfikować dwie klasy o tym samym prawdopodobieństwie (rozkład 50/50), maksymalną możliwą entropią jest Entropy = $-0,5 * \log_2(0,5) - 0,5 * \log_2(0,5) = 1,0$. Jednak gdy algorytm drzewa decyzyjnego wykryje funkcję, która może podzielić zestaw danych na dwie partycje, gdzie rozkład dwóch klas wynosi 40/60, średnia entropia informacyjna zmniejsza się:

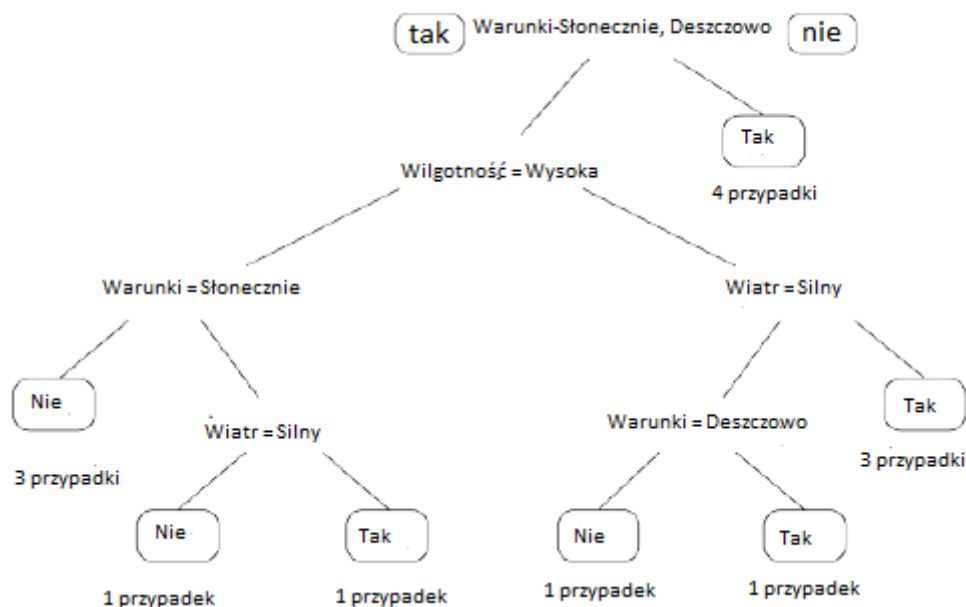
Entropia = $-0,4 * \log_2(0,4) - 0,6 * \log_2(0,6) = 0,97$

Zanotuj sumę entropii dla wszystkich klas. Przy zastosowaniu podziału 40/60 suma jest mniejsza niż teoretyczne maksimum 1 (pomniejszenie entropii). Pomyśl o entropii jako o pomiarze bałaganu w danych: im mniej bałaganu, tym więcej porządku i łatwiej odgadnąć właściwą klasę. Po pierwszym podziale algorytm próbuje dalej podzielić uzyskane partycje przy użyciu tej samej logiki zmniejszania entropii. Stopniowo dzieli każdą kolejną partycję danych, aż nie jest już możliwe dzielenie, ponieważ podpróbka jest pojedynczym przykładem lub ponieważ spełnia regułę zatrzymania. Reguły zatrzymywania są ograniczeniami ekspansji drzewa. Reguły te działają, biorąc pod uwagę trzy aspekty partycji: początkowy rozmiar partycji, wynikowy rozmiar partycji i informacje uzyskiwane przez podział. Reguły zatrzymywania są ważne, ponieważ algorytmy drzewa decyzyjnego przybliżają dużą liczbę funkcji; jednak szum i błędy danych mogą z łatwością wpływać na ten algorytm. W konsekwencji, w zależności od próby, niestabilność i wariancja uzyskanych szacunków wpływa na prognozy drzewa decyzyjnego.

Podjęcie decyzji na podstawie drzew

Jako przykład użycia drzewa decyzyjnego, w tej sekcji wykorzystano ten sam zestaw danych Ross Quinlan, który omówiono we wcześniejszej części. Korzystanie z tego zestawu danych pozwala nam przedstawić i opisać algorytm ID3, specjalny rodzaj drzewa decyzyjnego znalezione w artykule „Indukcja drzew decyzyjnych”, wspomniany wcześniej. Zbiór danych jest dość prosty, składa się tylko z 14 obserwacji w odniesieniu do warunków pogodowych, a wyniki wskazują, czy gra w tenisa jest odpowiednia. Przykład zawiera cztery funkcje: wygląd, temperaturę, wilgotność i wiatr, wszystkie

wyrażone za pomocą klas jakościowych zamiast pomiarów (można wyrazić liczbowo temperaturę, wilgotność i siłę wiatru), aby przekazać bardziej intuicyjne zrozumienie, w jaki sposób cechy pogody odnoszą się do wyniku. Po przetworzeniu tych funkcji przez algorytm można reprezentować zestaw danych za pomocą schematu drzewiastego, jak pokazano na rysunku



Jak pokazano na rysunku, można sprawdzić i odczytać zestaw reguł, dzieląc zestaw danych w celu utworzenia części w których prognozy są łatwiejsze, patrząc na najczęstszą klasę (w tym przypadku wynik, czyli grę w tenisa). Aby odczytać węzły drzewa, po prostu zacznij od najwyższego węzła, który odpowiada oryginalnym danym szkoleniowym; następnie zacznij czytać zasady. Zauważ, że każdy węzeł ma dwie pochodne: lewa gałąź oznacza, że górna reguła jest prawdziwa (podana jako tak w kwadratowym polu), a prawa oznacza, że jest fałszywa (podana jako nie w kwadratowym polu). Po prawej stronie pierwszej reguły widoczna jest ważna reguła końcowa (liść terminalny), w kółku, informująca o pozytywnym wyniku, Tak, który można odczytać jako grać w tenisa = prawda. Według tego węzła, gdy prognoza nie jest słoneczna (słońce) lub deszczowa (deszcz), można grać. (Liczby pod liściem terminalnym pokazują cztery przykłady potwierdzające tę regułę i zero zaprzeczające jej.) Zauważ, że możesz zrozumieć regułę lepiej, jeśli wynik po prostu stwierdził, że gdy prognoza jest zachmurzona, gra jest możliwa. Często reguły drzewa decyzyjnego nie są od razu używane i musisz je zinterpretować przed użyciem. Są jednak wyraźnie zrozumiałe (i znacznie lepsze niż wektor współczynników wartości). Po lewej drzewo postępuje zgodnie z innymi zasadami związanymi z wilgotnością. Ponownie, po lewej stronie, gdy wilgotność jest wysoka i słońce jest słoneczne, większość końcowych liści jest ujemna, z wyjątkiem sytuacji, gdy wiatr nie jest silny. Kiedy eksplorujesz gałęzie po prawej stronie, zobaczysz, że drzewo ujawnia, że gra jest zawsze możliwa, gdy wiatr nie jest silny lub gdy wiatr jest silny, ale nie pada.

Przycinanie zarośniętych drzew

Mimo że zestaw danych dotyczących gry w tenisa w poprzedniej sekcji ilustruje orzechy i piny drzewa decyzyjnego, ma niewielki potencjał probabilistyczny, ponieważ proponuje zestaw działań deterministycznych (nie ma sprzecznych instrukcji). Trening z prawdziwymi danymi zwykle nie obejmuje tak ostrych zasad, zapewniając w ten sposób miejsce na dwuznaczność i prawdopodobieństwo oczekiwanego wyniku. Drzewa decyzyjne mają w swoich szacunkach więcej wariacji niż stroniczości. Aby mniej dopasować dane, w przykładzie określono, że minimalny podział musi obejmować co najmniej pięć przykładów; przycina także drzewo. Przycinanie ma miejsce, gdy

drzewo jest w pełni dorosłe. Zaczynając od liści, przykład przycina drzewo gałęzi, wykazując niewielką poprawę w zmniejszaniu zdobywania informacji. Początkowo pozwalając drzewu się rozwijać, tolerowane są gałęzie z niewielką poprawą, ponieważ mogą odblokować bardziej interesujące gałęzie i liście. Prześledzenie od liści do korzenia i zachowanie tylko gałęzi, które mają pewną wartość predykcyjną, zmniejsza wariancję modelu, co powoduje, że wynikające z tego reguły są uciążliwe. W przypadku drzewa decyzyjnego przycinanie jest jak burza mózgów. Po pierwsze, kod generuje wszystkie możliwe konsekwencje drzewa (jak w przypadku pomysłów podczas burzy mózgów). Po drugie, gdy zakończy się burza mózgów, kod zachowuje tylko to, co naprawdę działa.