Notatki do wykładów: "Naiwny" klasyfikator Bayesowski

(c) Marcin Sydow

Naiwny Bayes

Tu zakładamy na ogół, że wszystkie atrybuty są kategoryczne. Mamy zbiór treningowy T składający się z N n-wymiarowych wektorów atrybutów.

Traktujemy atrybuty X_i i atrybut decyzyjny Y jako zmienne losowe

Mamy zaklasyfikować wektor $x = (x_1, x_2, ..., x_n)$

Stosujemy wzór Bayesa:

$$P(Y = y|X = x) = \frac{P(X = x|Y = y)P(Y = y)}{P(X = x)}$$

(interpretacja: prawdopobieństwo tego, że atrybut decyzyjny wynosi y pod warunkiem, że wartości atrybutów opisane są przez wektor x)

Zasada klasyfikatora Bayesa

Wektorowi x przydzielimy tę klasę (wartość atrybutu decyzyjnego) y, dla którego powyższe prawdopobieństwo jest najwyższe.

Obliczamy więc powyższe wyrażenie dla wszystkich możliwych klas (wartości atrybutu decyzyjnego Y) i wybieramy najwyższą wartośc prawdopobieństwa.

Ponieważ wszystkie powyższe porównywane wyrażenia mają ten sam mianownik ($P(X = x)$)), więc można go pominąć.

"Naiwny" klasyfikator Bayesa

Kluczowe dla "naiwnego" klasyfikatora Bayesowskiego jest ("naiwne") założenie, że atrybuty są parami niezależne, a więc:

$$P(X = (x_1, ..., x_n)|Y = y) = P(X_1 = x_1|Y = y)*...*P(X_n = x_n|Y = y)$$

Otrzymujemy więc po zastosowaniu powyższego założenia wzór:
$P(Y = y|X = (x_1, ..., x_n)) \propto P(X_1 = x_1|Y = y) * ... * P(X_n = x_n|Y = y) * P(Y = y)$
gdzie już bezpośrednio ze zbioru treningowego w prosty sposób można obliczyć oszacowania:

- $P(X_i = x_i|Y = y)$ (proporcja tych przypadków w zbiorze testowym, które mają wartość atrybutu $X_i = x_i$ wśród przypadków mających wartość atrybutu decycyjnego $Y = y$)
- oraz $P(Y = y)$ (proporcja przypadków w zbiorze treningowym, które mają wartość atrybutu decycyjnego $Y = y$)

Wygładzanie

Może się zdarzyć, że w zbiorze uczącym nie występuje żaden przypadek, w którym zachodzi $X_j = x_j$ oraz $Y = y$ dla pewnego atrybutu j.

W takim wypadku oszacowane prawdopobieństwo $P(X_i = x_i|Y = y)$ wynosiłoby zero i wyzerowało cały iloczyn, niezależnie od wartości pozostałych prawdopodobieństw $P(X_i = x_i|Y = y)$.

Aby tego uniknąć stosuje się tzw. wygładzanie, czyli zapewnienie, że zera zastępowane będą pewną (bardzo małą) wartością kosztem odpowiedniego zmniejszenia pozostałych (niezerowych) prawdopobieństw dla tego atrybutu.

Najprostsze wygładzanie

W najprostszym rodzaju wygładzania, do licznika proporcji dla danego atrybutu i dodajemy zawsze jeden a do mianownika tyle, ile jest różnych możliwych wartości tego atrybutu. W ten sposób zmodyfikowane prawdopodobieństwa sumują się do 1, ale nigdy nie wystąpi 0 nawet jak nie ma takiego przypadku w zbiorze treningowym.

Dziękuję za uwagę