Конспект лекции

А.Н. Шевляков

Машинное обучение

Вероятностные алгоритмы, байесовские классификаторы

# Вероятностные алгоритмы

Есть алгоритмы, которые выдают не метку класса, а вероятность принадлежности классам.

Примеры таких алгоритмов:

* наивный Байес
* модификация kNN
* модификация решающего дерева.

Зачем нужны такие алгоритмы предсказания?

* Когда данные будут подаваться на вход другим алгоритмам, вероятность несёт в себе дополнительную информацию: нашу уверенность в классификации.
* Возможность переложить ответственность на заказчика.

Как измерить качество работы вероятностного алгоритма?

Самый простой способ: округлить вероятности до целых значений и получить обычный алгоритм классификации. А потом для округленных значений считать precision, recall и т.д.

Такой способ не является оптимальным, особенно когда:

- классы не сбалансированы (число объектов одного класса много больше объектов другого класса)

- цены ошибок за неправильные классификации объектов класса 1 и класса 0 различны

- показатели качества работы вероятностного алгоритма должны учитывать величину порога округления.

# Наивный Байес

Как получить вероятности классов, когда признаков >1?

Пусть имеются данные:

Если применять формулу Байеса для объекта F, то нужно будет знать вероятности:

Pr(К’\*Л’ | М)

Pr(К’\*Л’ | Ж),

где

К={курит}, К’={не курит},

Л={любит кошек}, Л’= {не любит кошек},

\*= союз «и».

Как вычислить эти вероятности?

**Теорема о произведении вероятности**

Pr(A\*B)=Pr(A)\*Pr(B),

если события А и В не зависят друг от друга.

Решая нашу задачу, сделаем допущение, что курение НЕ ЗАВИСИТ от любви к кошкам и наоборот (а в жизни это действительно так?).

Тогда мы получаем:

Pr(К’\*Л’ | М)= Pr(К’ | М)\* Pr(Л’ | М),

Pr(К’\*Л’ | Ж)= Pr(К’ | Ж)\*Pr(Л’ | Ж),

а вероятности из правых частей равенств можно найти по таблице.

**Решение примера**

Из таблицы получаем вероятности:

Pr(Ж)=2/5 Pr(М)=3/5

Pr(К|Ж)=1/2 Pr(К’|Ж)=1/2

Pr(К|М)=2/3 Pr(К’|М)=1/3

Pr(Л|Ж)=1/2 Pr(Л’|Ж)=1/2

Pr(Л|М)=1/3 Pr(Л’|М)=2/3

По формуле Байеса для объекта F получаем вероятности принадлежности классам:

Pr(Ж|К’Л’)= Pr(К’Л’|Ж)\*Pr(Ж)/Pr(К’Л’)

Pr(М|К’Л’)= Pr(К’Л’|М)\*Pr(М)/Pr(К’Л’)

Используем предположение о независимости курения от любви к кошкам:

Pr(Ж|К’Л’)= Pr(К’|Ж)\* Pr(Л’|Ж)\* Pr(Ж)/Pr(К’Л’)

Pr(М|К’Л’)= Pr(К’|М)\* Pr(Л’|М)\* Pr(М)/Pr(К’Л’)

Подставляем числа:

Pr(Ж|К’Л’)= (1/2\* 1/3\* 2/5)/Pr(К’Л’)=(1/15)/Pr(К’Л’)

Pr(М|К’Л’)= (2/3\* 1/3\* 3/5)/Pr(К’Л’)=(2/15)/Pr(К’Л’)

По формуле Байеса для объекта F получаем вероятности принадлежности классам:

Pr(Ж|К’Л’)= Pr(К’Л’|Ж)\*Pr(Ж)/Pr(К’Л’)

Pr(М|К’Л’)= Pr(К’Л’|М)\*Pr(М)/Pr(К’Л’)

Используем предположение о независимости курения от любви к кошкам:

Pr(Ж|К’Л’)= Pr(К’|Ж)\* Pr(Л’|Ж)\* Pr(Ж)/Pr(К’Л’)

Pr(М|К’Л’)= Pr(К’|М)\* Pr(Л’|М)\* Pr(М)/Pr(К’Л’)

Подставляем числа:

Pr(Ж|К’Л’)= (1/2\* 1/3\* 2/5)/Pr(К’Л’)=(1/15)/Pr(К’Л’)

Pr(М|К’Л’)= (2/3\* 1/3\* 3/5)/Pr(К’Л’)=(2/15)/Pr(К’Л’)

Величину Pr(К’Л’) можно не вычислять, а воспользоваться равенством

Pr(Ж|К’Л’)+ Pr(М|К’Л’)=1

и получить Pr(Ж|К’Л’)=1/3 Pr(М|К’Л’)=2/3

Таким образом, принадлежность классу мужчин в 2 раза выше принадлежности классу женщин.

Pr(Ж|К’Л’)=1/3 Pr(М|К’Л’)=2/3

Аналогично обрабатываются таблицы с бОльшим числом признаков.

**Допущение о независимости признаков**

Работа классификатора существенно использует предположение о независимости признаков друг от друга.

Но независимость признаков (когда коэффициент корр≈0) не часто наблюдается на практике.

Поэтому этот классификатор называют наивным байесовским классификатором (НБК).

Когда между признаками существует сильная зависимость НБК может сильно ошибаться.

Плюсы НБК:

1. Простота и быстрая работа. Используется в системах массового обслуживания. Например, при фильтрации спама.
2. В некоторых случаях НБК хорошо работает даже для коррелированных признаков. Например, при фильтрации спама признаки, как правило, не являются независимыми.

Мы познакомились с алгоритмом «наивного Байеса». Теоретически он будет хорошо работать, если признаки не зависят друг от друга. Однако на практике он прекрасно работает и с зависимыми признаками.