



Наивный Байес



Наивный байесовский классификатор

Наивный байесовский классификатор — простой вероятностный классификатор, основанный на применении теоремы Байеса со строгими (наивными) предположениями о независимости.

- Задача классификации
- Обучение с учителем



Байесовская вероятностная модель

Имеем дата сет из $\mathbf{x}=(x_1, ..., x_n)$. Мы хотим знать вероятность \mathbf{x} отнести к классу y , y - переменная метки. Математически, нас интересует величина $P(y|\mathbf{x})$.

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

$$P(\mathbf{x}) = \sum_{y \in Y} P(y)P(\mathbf{x}|y)$$



Байесовская вероятностная модель

Нас интересует не точное значение $P(y|x)$, а отношение между вероятностями при разных y . К тому же, значение $P(x)$, как видно по формуле не зависит от y . Поэтому, знаменатель можно опустить.

Осталось вычислить величину $P(x|y)$, которая вычисляется сложно без дополнительных предположений.

$$\begin{aligned} P(\mathbf{x}|y) &= P(x_1, x_2, \dots, x_n|y) \\ &= P(x_1|y)P(x_2|x_1, y) \cdots P(x_n|x_1, x_2, \dots, x_{n-1}, y) \end{aligned}$$



Байесовская вероятностная модель

Наивное предположение Байеса — это, по сути, условное предположение независимости. Предполагается, что все функции в x взаимно независимы и зависят от категории. Математически мы предполагаем

$$P(x_i | x_1, x_2, \dots, x_{i-1}, y) = P(x_i | y)$$

$$\begin{aligned} P(\mathbf{x} | y) &= P(x_1, x_2, \dots, x_n | y) \\ &= \prod_{i=1}^n P(x_i | y) \end{aligned}$$



Байесовская вероятностная модель

Итоговый вид нашей формулы:

$$P(y|\mathbf{x}) = \frac{P(y) \prod_{i=1}^n P(x_i|y)}{P(\mathbf{x})}$$
$$\propto P(y) \prod_{i=1}^n P(x_i|y)$$

$$\tilde{y} = \operatorname{argmax}_y P(y) \prod_{i=1}^n P(x_i|y)$$



Достоинства

- Алгоритм легко и быстро предсказывает класс тестового набора данных. Он также хорошо справляется с многоклассовым прогнозированием.
- Производительность наивного байесовского классификатора лучше, чем у других простых алгоритмов, таких как логистическая регрессия. Более того, вам требуется меньше обучающих данных.
- Он хорошо работает с категориальными признаками(по сравнению с числовыми). Для числовых признаков предполагается нормальное распределение, что может быть серьезным допущением в точности нашего алгоритма.
- Легок в реализации



Недостатки

- Если переменная имеет категорию (в тестовом наборе данных), которая не наблюдалась в обучающем наборе данных, то модель присвоит 0 (нулевую) вероятность и не сможет сделать предсказание. Это часто называют нулевой частотой. Чтобы решить эту проблему, мы можем использовать технику сглаживания. Один из самых простых методов сглаживания называется оценкой Лапласа.
- Ограничением данного алгоритма является предположение о независимости признаков. Однако в реальных задачах полностью независимые признаки встречаются крайне редко.