Naive Bayes Classifier

Quách Đình Hoàng

2022/10/06

Contents

Introduction

Naive Bayes classifier (NBC)

Introduction

Probability and Machine Learning

- Supervised learning: We want predict y from x.
- A way to do this task is to use the idea from probability
  - Determine $p(y|x)$ - discriminative model.
  - Determine $p(x, y)$ - generative model
- Probabilistic machine learning is an important branch in machine learning.
- When y is a categorical variable we have a classification problem
- When y is a numerical variable we have a regression problem

Bayes theorem

$$P(h|D) = \frac{P(D|h)P(h)}{P(D)}$$

- $P(h)$ - prior probability: The probability of hypothesis $h \rightarrow$ prior probability
- $P(D)$ - evidence: The probability of data, D
- $P(D|h)$ - likelihood: The probability of, D, given hypothesis, h
- $P(h|D)$ - posterior: The probability of hypothesis, h, given data, D

Maximum likelihood estimation (MLE)

- Given a set of hypotheses, H. Find the hypothesis that maximize the likelihood, $P(D|h)$

$$h_{ML} = \underset{h}{\operatorname{argmax}} P(D|h)$$

Maximum a posteriori (MAP)

- Given a set of hypotheses, H. Find the hypothesis that maximize the posterior, $P(h|D)$

$$
\begin{aligned}
h_{MAP} &= \underset{h \in H}{\operatorname{argmax}} P(h|D) \\
&= \underset{h \in H}{\operatorname{argmax}} \frac{P(D|h)P(h)}{P(D)} \quad (Bayes theorem) \\
&= \underset{h \in H}{\operatorname{argmax}} P(D|h)P(h) \quad (P(D) \text{ is independent of } h)
\end{aligned}
$$

Naive Bayes classifier (NBC)

Naive Bayes classifier (NBC)

- Given
  - A training set D, has training instances $x = (x_1, x_2, \cdots, x_d)$
  - A set of predefined labels/classes: $C = \{c_1, c_2, \cdots, c_m\}$
- With a new instance, z, NBC want to find the label/class with the greatest probability given z

$$
\begin{aligned}
c_{MAP} &= \underset{c_i \in C}{\operatorname{argmax}} P(c_i|z) \\
&= \underset{c_i \in C}{\operatorname{argmax}} \frac{P(z|c_i)P(c_i)}{P(z)} \quad (Bayes theorem) \\
&= \underset{c_i \in C}{\operatorname{argmax}} P(z|c_i)P(c_i) \quad (P(z) \text{ is independent of } c_i)
\end{aligned}
$$

Naive Bayes classifier (NBC)

- This is a probabilistic classification, use Bayes' theorem to find the label/class with the greatest probability given z

$$c_{MAP} = \underset{c_i \in C}{\operatorname{argmax}} P(z|c_i)P(c_i) = \underset{c_i \in C}{\operatorname{argmax}} P(z_1, z_2, \cdots, z_d|c_i)P(c_i)$$

- This model is based on the naive) assumption that the features are conditionally independent given class label.

$$P(z|c_i) = P(z_1, z_2, \cdots, z_d|c_i) = \prod_{j=1}^{d} p(X_j = z_j|Y = c_i)$$

- Although this assumption may not be true, the Naive Bayes classifier usually gives good results in practice compared to other more complex methods.

Naive Bayes Training

X_i is a categorical variable

- Using MLE Estimate

$$P(X_j = z_j | Y = c_i) = \frac{count(X_j = z_j, Y = c_i)}{count(Y = c_i)}$$

- Using Laplace MAP estimate

$$p(X_j = z_j | Y = c_i) = \frac{count(Z_j = z_j, Y = c_i) + 1}{count(Y = c_i) + v}$$

where: v is the total number of values that X_j can take.

Naive Bayes Training

X_j is a numertical variable

- Discretized
  - Partition the range of values of X_j into intervals and replace the continuous value with the value representing the corresponding interval.
- Use the probability density function
  - Assume X_j follows a distribution (e.g. normal distribution)
  - Use training data to estimate the parameters of the distribution (e.g. μ, σ with a normal distribution)
  - Calculate the probability $P(X_j = z_j | Y = c_i)$ based on the estimated distribution

Naive Bayes Training

X_j is a numertical variable

- If X_j is continuous, $P(X_j = z_j | Y = c_i) = 0$
- However, we can estimate $P(z_j - \epsilon \leq X_j \leq z_j + \epsilon | Y = c_i)$
- Assumptions

$$P(X_j | Y_i) = \frac{1}{\sqrt{2\pi\sigma_{ji}^2}} e^{-\frac{(X_i - \mu_{ji})^2}{2\sigma_{ji}^2}}$$

- Then

$$\begin{aligned} P(z_j - \epsilon \leq X_j \leq z_j + \epsilon | Y = c_i) &= \int_{z_j - \epsilon}^{z_j + \epsilon} \frac{1}{\sqrt{2\pi\sigma_{ji}^2}} e^{-\frac{(z - \mu_{ji})^2}{2\sigma_{ji}^2}} z dz \\ &\approx 2\epsilon f(z_j, \mu_{ji}, \sigma_{ji}) \end{aligned}$$

Naive Bayes Classifier Summary

- NBC's training
  - For each label/class $c_i \in C$
    - Estimate priori probability: $P(c_i)$
    - For each attribute value z_j, estimate $P(X_j = z_j | Y = c_i)$
- NBC's prediction
  - For each label/class $c_i \in C$, compute

$$P(c_i) \prod_{j=1}^{d} p(X_j = z_j | Y = c_i)$$

  - Choose label/class $c_i \in C$ with the highest probability

$$c_{MAP} = \underset{c_i \in C}{\mathrm{argmax}} P(c_i) \prod_{j=1}^{d} p(X_j = z_j | Y = c_i)$$