How to Keep your Customers

An Introduction to Naïve Bayes Classifiers

Lairoongroj Jinnawat (Jake)

BA2203 Presentation

24 October 2022

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

1. Naïve Bayes Classifiers
2. Example: Predicting Lapse

Recap: Bayes' Theorem

Theorem (Bayes' Theorem)

Given events A and B where $P(B) \neq 0$, we have

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}.$$

- Very powerful!

Naïve Bayes Classifiers: Definition

Model from machine learning (Hastie et al., 2009).

- Outcomes or classes: $C_1, C_2, \ldots$
- Observed predictor variables or features: $\mathbf{x} = (x_1, x_2, \ldots, x_n)$
- Assume all features are mutually independent.

Naïve Bayes Classifiers: Definition

Model from machine learning (Hastie et al., 2009).

- Outcomes or classes: $C_1, C_2, \ldots$
- Observed predictor variables or features: $\mathbf{x} = (x_1, x_2, \ldots, x_n)$
- Assume all features are mutually independent.

Theorem

$$P(C_k | x_1, \ldots, x_n) \propto P(C_k) \prod_{i=1}^{n} P(x_i | C_k).$$

Naïve Bayes Classifiers: Definition

Proof.

Consider the event of observing C_k *and* features $x_1, x_2, \ldots$. Extracting the x_i's, we have

$$
\begin{aligned}
P(C_k, x_1, \ldots, x_n) &= P(x_1 | C_k, x_2, \ldots, x_n) P(C_k, x_2, \ldots, x_n) \\
&= P(x_1 | C_k, x_2, \ldots, x_n) P(x_2 | C_k, x_3, \ldots, x_n) P(C_k, x_3, \ldots, x_n) \\
&= \ldots \\
&= P(x_1 | C_k, \cancel{x_2, \ldots, x_n}) P(x_2 | C_k, \cancel{x_3, \ldots, x_n}) \cdots P(C_k) \\
&= P(C_k) \prod_{i=1}^{n} P(x_i | C_k).
\end{aligned}
$$

$$
\therefore P(C_k | x_1, \ldots, x_n) = \frac{P(C_k, x_1, \ldots, x_n)}{P(x_1, \ldots, x_n)} \propto P(C_k) \prod_{i=1}^{n} P(x_i | C_k). \quad \square
$$

Naïve Bayes Classifiers: Definition

- Decision-making: Classify observation as C_k which maximises $P(C_k)\prod_{i=1}^{n} P(x_i|C_k)$.
- Probabilities are obtained from past observations, i.e., training data.
- Note: Features can be discrete (multinomial, Bernoulli) or continuous (normal, nonparametric) (John and Langley, 1995).

A Basic Life Insurance Product

You are an actuarial analyst at a life insurer, looking to create a simple model for whether an individual policy will *lapse in a given year* (0 or 1) and identify ways to retain customers.

Looking through the literature (Fang and Kung, 2021; Eling and Kochanski, 2013), you decide on the following parameters/assumptions:

A Basic Life Insurance Product

You are an actuarial analyst at a life insurer, looking to create a simple model for whether an individual policy will *lapse in a given year* (0 or 1) and identify ways to retain customers.

Looking through the literature (Fang and Kung, 2021; Eling and Kochanski, 2013), you decide on the following parameters/assumptions:

- Classes: policy lapsed ($= 1$) or policy did not lapse ($= 0$)
- Features:
  - age band (young, middle-aged, old);
  - gender (male, female);
  - smoker status (smoker, non-smoker); and
  - macroeconomic conditions (good, bad).
- Assume all features are mutually independent.

One Class: Age Band

Suppose we wish to predict whether a middle-aged policyholder will lapse.

Lapsed? \ Age Band	Young	Middle-Aged	Old	Total
Yes	14	45	49	108
No	379	315	198	892
Total	392	364	244	1000

One Class: Age Band

Suppose we wish to predict whether a middle-aged policyholder will lapse.

Lapsed? \ Age Band	Young	Middle-Aged	Old	Total
Yes	14	45	49	108
No	379	315	198	892
Total	392	364	244	1000

$$P(\text{middle-aged} \mid 1) = \frac{45}{108} \approx 0.435. \tag{1}$$

$$P(\text{middle-aged} \mid 0) = \frac{315}{892} \approx 0.353. \tag{2}$$

All Classes

Now suppose that the policyholder is a middle-aged male smoker, and economic conditions that year are bad. Let the vector of features x_i be $\mathbf{x}$.

Features / Conditional Prob.	Middle-Aged	Male	Smoker	Bad Cond$^{\underline{n}}$s
$P(x_i \mid 1)$	0.4167	0.6204	0.7315	0.7130
$P(x_i \mid 0)$	0.3531	0.4664	0.3094	0.2365

$$
\begin{aligned}
P(1 \mid \mathbf{x}) &\propto P(1) \prod_{x_i \in \mathbf{x}} P(x_i \mid 1) \\
&= \frac{108}{1000}(0.4167)(0.6204)(0.7315)(0.7130) \qquad (3) \\
&\approx 1.6\%.
\end{aligned}
$$

All Classes

Similarly,

$$
\begin{aligned}
P(0 \mid \mathbf{x}) &\propto P(0) \prod_{x_i \in \mathbf{x}} P(x_i \mid 0) \\
&= \frac{892}{1000}(0.3531)(0.4664)(0.3094)(0.2365) \\
&\approx 1.1\%.
\end{aligned} \tag{4}
$$

Since $P(1 \mid \mathbf{x}) > P(0 \mid \mathbf{x})$, classify the policyholder as likely to lapse.

Real-World Applications

- Possible *interventions* to recommend: Improve health status, e.g., offer discounts for quitting smoking.
- In reality, many more features may be used (with appropriate penalisation). Large sample size, metrics like *sensitivity* and *specificity* for comparing models.

Real-World Applications

- Possible *interventions* to recommend: Improve health status, e.g., offer discounts for quitting smoking.
- In reality, many more features may be used (with appropriate penalisation). Large sample size, metrics like *sensitivity* and *specificity* for comparing models.
- Other uses:
  - document classification (spam filtering) (Sahami et al., 1998),
  - sentiment analysis (Pang et al., 2002),
  - medical predictions (Khanna and Sharma, 2018), etc.

Real-World Applications

- Possible *interventions* to recommend: Improve health status, e.g., offer discounts for quitting smoking.
- In reality, many more features may be used (with appropriate penalisation). Large sample size, metrics like *sensitivity* and *specificity* for comparing models.
- Other uses:
  - document classification (spam filtering) (Sahami et al., 1998),
  - sentiment analysis (Pang et al., 2002),
  - medical predictions (Khanna and Sharma, 2018), etc.
- Naïve Bayes classifiers are simple, fast, and easy to implement (especially with Python packages like `scikit-learn`).

Thank you!

Contact me!

Slides

Eling, M., & Kochanski, M. (2013). Research on lapse in life insurance: What has been done and what needs to be done? *The Journal of Risk Finance*, *14*(4), 392–413. https://doi.org/10.1108/JRF-12-2012-0088

Fang, H., & Kung, E. (2021). Why do life insurance policyholders lapse? the roles of income, health, and bequest motive shocks. *Journal of Risk and Insurance*, *88*(4), 937–970. https://doi.org/10.1111/jori.12332

Hastie, T., Tibshirani, R., & Friedman, J. (2009). Kernel smoothing methods. In T. Hastie, R. Tibshirani, & J. Friedman (Eds.), *The elements of statistical learning: Data mining, inference, and prediction* (pp. 191–218). Springer. https://doi.org/10.1007/978-0-387-84858-7_6

John, G. H., & Langley, P. (1995). Estimating continuous distributions in bayesian classifiers. *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*, 338–345.

Khanna, D., & Sharma, A. (2018). Kernel-based naive bayes classifier for medical predictions. In V. Bhateja, C. A. Coello Coello, S. C. Satapathy, & P. K. Pattnaik (Eds.), *Intelligent engineering informatics* (pp. 91–101). Springer. https://doi.org/10.1007/978-981-10-7566-7_10

Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? sentiment classification using machine learning techniques. *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP 2002)*, 79–86. https://doi.org/10.3115/1118693.1118704

Sahami, M., Dumais, S., Heckerman, D., & Horvitz, E. (1998). A bayesian approach to filtering junk e-mail.