Tut 5: Naive Bayes Classifier

Jan 2023

The general mathematical formulation of a generative model:

$$\begin{aligned} h_D(\boldsymbol{x}) &= \underset{j\in\{1,\ldots,K\}}{\operatorname{argmax}}\ \mathbb{P}(Y=j|\boldsymbol{X}=\boldsymbol{x}) \\ &= \underset{j\in\{1,\ldots,K\}}{\operatorname{argmax}}\ \frac{\mathbb{P}(\boldsymbol{X}=\boldsymbol{x}|Y=j)\mathbb{P}(Y=j)}{\mathbb{P}(\boldsymbol{X}=\boldsymbol{x})} \\ &= \underset{j\in\{1,\ldots,K\}}{\operatorname{argmax}}\ \mathbb{P}(\boldsymbol{X}=\boldsymbol{x}|Y=j)\mathbb{P}(Y=j) \\ &= \underset{j\in\{1,\ldots,K\}}{\operatorname{argmax}}\ [\ln\mathbb{P}(\boldsymbol{X}=\boldsymbol{x}|Y=j)+\ln\mathbb{P}(Y=j)] \end{aligned} \tag{5.1}$$

Naive Bayes:

$$\boxed{\mathbb{P}(\boldsymbol{X}=\boldsymbol{x}|Y=j)\approx\prod_{i=1}^{p}\mathbb{P}(X_i=x_i|Y=j)}$$

1. (Jan 2022 Final Q4(a)) The training data for part (a) is given in Table 4.1.

Table 4.1: Training data for credit card application approval.

Age	PriorDefault	Employed	Approved
59.67	Yes	False	+
27.25	No	True	-
20.67	No	False	-
16.50	No	False	-
26.67	Yes	True	+
37.50	Yes	False	-
36.25	Yes	True	+
21.17	No	False	-
32.33	Yes	False	+
58.42	Yes	True	+

Use the Naïve Bayes classifier model without Laplace smoothing to predict if the credit card approval is positive or negative for the person is of age 38.17, has a prior default and is employed. (10 marks)

Solution. Let Y=Approved, X_1=Age, X_2=PriorDefault, X_3=Employed.

$$\begin{aligned} &P(Y=+|X_1=38.17,X_2=Yes,X_3=True) \\ \propto &P(X_1=38.17|Y=+)\times P(X_2=Yes|Y=+)\times P(X_3=True|Y=+)P(Y=+) \end{aligned}$$ [1 mark]

Y	$P(Y)$	$X_1=38.17$	$X_2=Yes$	$X_3=True$	Product	Prob
$+$	$\frac{5}{10}=0.5$	0.02491317	$\frac{5}{5}=1$	$\frac{3}{5}=0.6$	0.0074740	0.9681
$-$	$\frac{5}{10}=0.5$	0.01230699	$\frac{1}{5}=0.2$	$\frac{1}{5}=0.2$	0.0002461	0.0319
	[1.5 marks]	[3 marks]	[1.5 marks]	[1.5 marks]	[0.5 mark]	

Using scientific calculator, we can obtain the estimate:

$$\mu_+ = \frac{59.67 + 26.67 + 36.25 + 32.33 + 58.42}{5} = 42.668$$
$$\sigma_+ = \sqrt{\frac{(59.67 - \mu_+)^2 + (26.67 - \mu_+)^2 + ... + (58.42 - \mu_+)^2}{5-1}} = 15.33945$$

$$P(X_1 = 38.17|Y = +) = \frac{1}{\sqrt{2\pi}(15.33945)} \exp(-\frac{(38.17 - 42.668)^2}{2(235.2986)}) = 0.02491317$$

Similarly,

$$\mu_- = 24.618$$
$$\sigma_- = 8.158544805$$

Since the product $P(X_1 = 38.17|Y = +) \times P(X_2 = Yes|Y = +) \times P(X_3 = True|Y = +)P(Y = +) > P(X_1 = 38.17|Y = -) \times P(X_2 = Yes|Y = -) \times P(X_3 = True|Y = -)P(Y = -)$, the credit card approval is **positive**. . . . .[1 mark] □

2. Ahmad would like to construct a model to decide if a day is suitable to play tennis. The table below shows the results whether to play tennis, based on Outlook, Temperature and Wind, collected by Ahmad.

Day	Outlook	Temperature	Wind	PlayTennis
D1	Sunny	34	Weak	No
D2	Sunny	32	Strong	No
D3	Overcast	28	Weak	Yes
D4	Rain	22	Weak	Yes
D5	Rain	16	Weak	Yes
D6	Rain	8	Strong	No
D7	Overcast	12	Strong	Yes
D8	Sunny	20	Weak	No
D9	Sunny	10	Weak	Yes
D10	Rain	23	Weak	Yes
D11	Sunny	19	Strong	Yes
D12	Overcast	21	Strong	Yes
D13	Overcast	31	Weak	Yes
D14	Rain	25	Strong	No

Using Naïve Bayes approach with Laplace smoothing, predict whether a sunny day with strong wind, 27°C, is suitable to play tennis.

Solution. Let $y = PlayTennis(Yes = 1; No = 0)$

X_1=Outlook; X_2=Temperature; X_3=Wind

New observation: $x_1^* = sunny$; $x_2^* = 27$; $x_3^* = strong$

Steps for finding the posterior $\mathbb{P}(Y = 1|\boldsymbol{X} = \boldsymbol{x}^*)$.

- Prior, $\mathbb{P}(Y = 1) = \frac{9}{14}$
- Density functions,

  $\mathbb{P}(X_1 = sunny|Y = 1) = \frac{2+1}{9+3} = \frac{1}{4}$

$$\mathbb{P}(X_2 = 27|Y=1) = \frac{1}{\sqrt{2\pi(s^2_{x_{2:y=1}})}} e^{-\frac{(x_2^* - \overline{x_{2:y=1}})^2}{2s^2_{x_{2:y=1}}}} = \frac{1}{\sqrt{2\pi}(6.8880)} e^{-\frac{(27-20.2222)^2}{2(47.4445)}} = 0.0357$$

where $\overline{x_{2:y=1}} = 20.2222$; $s_{x_{2:y=1}} = 6.8880$

$$\mathbb{P}(X_3 = strong|Y=1) = \frac{3+1}{9+2} = \frac{4}{11}$$

- Hence, posterior probability for PlayTennis=Yes is

$$\begin{aligned}
&\mathbb{P}(\hat{Y} = 1|\boldsymbol{X} = \boldsymbol{x}^*)\\
&\propto P(Y=1)\cdot\mathbb{P}(X_1 = sunny|Y=1)\cdot\mathbb{P}(X_2 = 27|Y=1)\cdot\mathbb{P}(X_3 = strong|y=1)\\
&= \frac{9}{14}\cdot\frac{1}{4}\cdot 0.0357\cdot\frac{4}{11} \approx 0.0021
\end{aligned}$$

Steps for finding the posterior $\mathbb{P}(Y=0|\boldsymbol{X} = \boldsymbol{x}^*)$.

- Prior, $\mathbb{P}(Y=0) = \dfrac{5}{14}$
- Density functions,

$$\mathbb{P}(X_1 = sunny|Y=0) = \frac{3+1}{5+3} = \frac{1}{2}$$

$$\mathbb{P}(X_2 = 27|Y=0) = \frac{1}{\sqrt{2\pi}(s^2_{x_{2:y=0}})} e^{-\frac{(x_2^* - \overline{x_{2:y=0}})^2}{2s^2_{x_{2:y=0}}}} = \frac{1}{\sqrt{2\pi}(10.4499)} e^{-\frac{(27-23.8)^2}{2(109.20)^2}} = 0.0364$$

where $\overline{x_{2:y=0}} = 23.8000$; $s_{x_{2:y=0}} = 10.4499$

$\mathbb{P}(X_3 = strong|y=0) = \frac{3+1}{5+2} = \frac{4}{7}$

Hence, posterior probability for (PlayTennis = No) is

$$\begin{aligned}
&\mathbb{P}(Y = 0|\boldsymbol{X} = \boldsymbol{x}^*)\\
&\propto \mathbb{P}(y=0)\cdot\mathbb{P}(X_1 = sunny|Y=0)\cdot\mathbb{P}(X_2 = 27|Y=0)\cdot\mathbb{P}(X_3 = strong|Y=0)\\
&= \frac{5}{14}\cdot\frac{1}{2}\cdot 0.0364\cdot\frac{4}{7} \approx 0.0037
\end{aligned}$$

Since $\mathbb{P}(Y=0|\boldsymbol{X} = \boldsymbol{x}^*) > \mathbb{P}(Y=1|\boldsymbol{X} = \boldsymbol{x}^*)$, the day is not suitable to play tennis. □

3. (Jan 2021 Final Q4(b)) Suppose the mood (M) of a student is affected by two features, the weather (W) and his result (R) and the Table 4.2.

Table 4.2: Observed Data.

Weather (W)	Result (R)	Mood (M)
Bad	Poor	Unhappy
Good	Poor	Unhappy
Good	Poor	Unhappy
Good	Poor	Unhappy
Bad	Good	Unhappy
Bad	Good	Happy
Bad	Good	Happy
Good	Good	Happy

(a) Using Table 4.2 and a Naive Bayes classifier to predict the mood if today's situation is that the weather is good, the result is good. Show your computations clearly and write down the classifier's prediction. (1.5 marks)

Solution. Let Unhappy=U, Happy=H, G=Good. Then

$$P(M = U|W = G, R = G)$$
$$\propto P(W = G|M = U) \times P(R = G|M = U) \times P(M = U) = \frac{3}{5} \times \frac{1}{5} \times \frac{5}{8} = 0.075$$
[0.6 mark]

$$P(M = H|W = G, R = G)$$
$$\propto P(W = G|M = H) \times P(R = G|M = H) \times P(M = H) = \frac{1}{3} \times \frac{3}{3} \times \frac{3}{8} = 0.125$$
[0.6 mark]

The classifier's prediction of the mood is **Happy**. ....[0.3 mark] □

(b) Using Table 4.2 and a Naive Bayes classifier to predict the mood if today's situation is that the weather is bad, the result is poor. Show your computations clearly and write down the classifier's prediction. (1.5 marks)

Solution. Let Unhappy=U, Happy=H, B=Bad, P=Poor. Then

$$P(M = U|W = B, R = P)$$
$$\propto P(W = B|M = U) \times P(R = P|M = U) \times P(M = U) = \frac{2}{5} \times \frac{4}{5} \times \frac{5}{8} = 0.2$$
[0.6 mark]

$$P(M = Happy|W = R, R = P)$$
$$\propto P(W = B|M = H) \times P(R = P|M = H) \times P(M = H) = \frac{2}{3} \times \frac{0}{3} \times \frac{3}{8} = 0$$
[0.6 mark]

The classifier's prediction of the mood is **Unhappy**. .... [0.3 mark] □

(c) Suppose an additional feature, exercise (E), which indicates that the student will carry out outdoor exercise or not, is added to the Table 4.2 to form Table 4.3.

Table 4.3: Observed Data with New Feature.

Weather (W)	Result (R)	Exercise (E)	Mood (M)
Bad	Poor	No	Unhappy
Good	Poor	Yes	Unhappy
Good	Poor	Yes	Unhappy
Good	Poor	Yes	Unhappy
Bad	Good	No	Unhappy
Bad	Good	No	Happy
Bad	Good	No	Happy
Good	Good	Yes	Happy

Using Table 4.3 and the Naive Bayes Classifier to the mood if W=Good, R= Good, E=Yes. Show your computations and the classifier's prediction. Will the new feature improve the performance of the Naive Bayes classifier from the one built based on Table 4.2? Justify your answer. (2 marks)

Solution. Let Unhappy=U, Happy=H, G=Good, Y=Yes. Then

$$P(M = U|W = G, R = G, E = Y)$$
$$\propto P(W = G|M = U) \times P(R = G|M = U) \times P(E = Y|M = U) \times P(M = U)$$
$$= \frac{3}{5} \times \frac{1}{5} \times \frac{3}{5} \times \frac{5}{8} = 0.045$$
[0.3 mark]

$$P(M = H|W = G, R = G, E = Y)$$
$$\propto P(W = G|M = H) \times P(R = G|M = H) \times P(E = Y|M = H) \times P(G = H)$$

$= \frac{1}{3} \times \frac{3}{3} \times \frac{1}{3} \times \frac{3}{8} = 0.04166667$ [0.3 mark]

The classifier's prediction of the mood is **Unhappy**. .... [0.2 mark]

No. .... [0.2 mark]

The new feature E will not improve the performance of the Naive Bayes classifier's prediction because the new feature E is correlated with the feature W and violates the assumption in Naive Bayes classifier. .... [1 mark] □

4. (Final Assessment May 2020 Q2) The testing dataset of an insurance claim is given in Table 2.1. The variables "gender", "bmi", "age_bracket" and "previous_claim" are the predictors and the "claim" is the response.

Table 2.1: The testing data of an insurance claim (randomly sampled with repeated entry).

gender	bmi	age_bracket	previous_claim	claim
female	under_weight	18-30	0	no_claim
female	under_weight	18-30	0	no_claim
male	over_weight	31-50	0	no_claim
female	under_weight	50+	1	no_claim
male	normal_weight	18-30	0	no_claim
female	under_weight	18-30	1	no_claim
male	over_weight	18-30	1	no_claim
male	over_weight	50+	1	claim
female	normal_weight	18-30	0	no_claim
female	obese	50+	0	claim

The "gender" is binary categorical data, the "bmi" is a four-value categorical data with values under_weight, normal_weight, over_weight and obese, the "age_bracket" is a three-value categorical data with value "18-30", "31-50" and "50+", the "previous_claim" is a binary categorical data with 0 indicating "no previous claim" and 1 indicating "having a previous claim". The "claim" is a binary response with values "no_claim" (negative class, with value 1) and "claim" (positive class, with value 0).

(b) Write down the mathematical formula for the Naive Bayes model with the predictors and response in Table 2.3. Use the Naive Bayes model trained on the training data from Table 2.3 to **predict** the "claim" of the insurance data in Table 2.1 as well as **evaluating** the performance of the model by calculating the confusion matrix, accuracy, sensitivity, specificity, PPV, NPV of the Naive Bayes model.

Table 2.3: The training dataset of an insurance claim data for Naive Bayes model.

Obs.	gender	bmi	age_bracket	previous_claim	claim
1	female	obese	50+	1	no_claim
2	female	under_weight	31-50	0	no_claim
3	male	under_weight	31-50	1	no_claim
4	female	over_weight	18-30	1	no_claim
5	female	normal_weight	31-50	0	no_claim
6	female	under_weight	31-50	0	no_claim
7	female	obese	18-30	0	no_claim
8	male	under_weight	50+	1	no_claim
9	female	normal_weight	31-50	0	no_claim
10	male	over_weight	31-50	0	no_claim
11	female	normal_weight	50+	0	claim
12	male	over_weight	31-50	1	claim
13	male	under_weight	31-50	1	claim
14	male	over_weight	31-50	1	claim
15	male	obese	50+	0	claim
16	male	under_weight	50+	0	claim
17	female	obese	31-50	1	claim
18	female	under_weight	50+	1	claim
19	female	normal_weight	50+	1	claim
20	female	under_weight	18-30	1	claim

Note: The default cut-off is 0.5.

Solution. Let X be the predictors; g be the predictor "gender" with F (female) and M (male); b be the predictor "bmi" with UW (under weight), OW (over weight), NW (normal weight), OB (obese); a be the predictor "age bracket" with a18 (18-30), a31 (31-50) and a50 (50+); p be the predictor "previous claim"; Y be the "actual" response "claim". The Naive Bayes model is

$$\mathbb{P}(Y|X) \propto \mathbb{P}(Y) \cdot \mathbb{P}(g|Y) \cdot \mathbb{P}(b|Y) \cdot \mathbb{P}(a|Y) \cdot \mathbb{P}(p|Y) = \text{prop}. \qquad [0.5 \text{ mark}]$$

Let $\widehat{Y}$ be the predicted response. Note that in the question, "no_claim" has a value 1 (negative) and "claim" has a value 0 (positive) which we will follow here. For the given training data, we have

$$\mathbb{P}(Y = 1) = \mathbb{P}(Y = 0) = \frac{10}{20} = 0.5. \qquad [0.5 \text{ mark}]$$

Since it will not contribute to our calculation, we can actually ignore it. However, it will be maintained to match textbook algorithm.

From Table 2.1, we need to calculate

$\mathbb{P}(g = F\|Y = 1) = 0.7$	$\mathbb{P}(g = M\|Y = 1) = 0.3$	
$\mathbb{P}(g = F\|Y = 0) = 0.5$	$\mathbb{P}(g = M\|Y = 0) = 0.5$	
$\mathbb{P}(b = UW\|Y = 1) = 0.4$	$\mathbb{P}(b = NW\|Y = 1) = 0.2$	
$\mathbb{P}(b = OW\|Y = 1) = 0.2$	$\mathbb{P}(b = OB\|Y = 1) = 0.2$	
$\mathbb{P}(b = UW\|Y = 0) = 0.4$	$\mathbb{P}(b = NW\|Y = 0) = 0.2$	
$\mathbb{P}(b = OW\|Y = 0) = 0.2$	$\mathbb{P}(b = OB\|Y = 0) = 0.2$	
$\mathbb{P}(a = a18\|Y = 1) = 0.2$	$\mathbb{P}(a = a31\|Y = 1) = 0.6$	$\mathbb{P}(a = a50\|Y = 1) = 0.2$
$\mathbb{P}(a = a18\|Y = 0) = 0.1$	$\mathbb{P}(a = a31\|Y = 0) = 0.4$	$\mathbb{P}(a = a50\|Y = 0) = 0.5$
$\mathbb{P}(p = 1\|Y = 1) = 0.4$	$\mathbb{P}(p = 0\|Y = 1) = 0.6$	
$\mathbb{P}(p = 1\|Y = 0) = 0.7$	$\mathbb{P}(p = 0\|Y = 0) = 0.3$	

prior	$\mathbb{P}(g\|Y)$	$\mathbb{P}(b\|Y)$	$\mathbb{P}(a\|Y)$	$\mathbb{P}(p\|Y)$	prop	$\widehat{Y}$	Y
$\mathbb{P}(Y=1)=0.5$	0.7	0.4	0.2	0.6	0.0168	✓	no_claim
$\mathbb{P}(Y=0)=0.5$	0.5	0.4	0.1	0.3	0.0030		
$\mathbb{P}(Y=1)=0.5$	0.7	0.4	0.2	0.6	0.0168	✓	no_claim
$\mathbb{P}(Y=0)=0.5$	0.5	0.4	0.1	0.3	0.0030		
$\mathbb{P}(Y=1)=0.5$	0.3	0.2	0.6	0.6	0.0108	✓	no_claim
$\mathbb{P}(Y=0)=0.5$	0.5	0.2	0.4	0.3	0.0060		
$\mathbb{P}(Y=1)=0.5$	0.7	0.4	0.2	0.4	0.0112		no_claim
$\mathbb{P}(Y=0)=0.5$	0.5	0.4	0.5	0.7	0.0350	✓	
$\mathbb{P}(Y=1)=0.5$	0.3	0.2	0.2	0.6	0.0036	✓	no_claim
$\mathbb{P}(Y=0)=0.5$	0.5	0.2	0.1	0.3	0.0015		
$\mathbb{P}(Y=1)=0.5$	0.7	0.4	0.2	0.4	0.0112	✓	no_claim
$\mathbb{P}(Y=0)=0.5$	0.5	0.4	0.1	0.7	0.0070		
$\mathbb{P}(Y=1)=0.5$	0.3	0.2	0.2	0.4	0.0024		no_claim
$\mathbb{P}(Y=0)=0.5$	0.5	0.2	0.1	0.7	0.0035	✓	
$\mathbb{P}(Y=1)=0.5$	0.3	0.2	0.2	0.4	0.0024		
$\mathbb{P}(Y=0)=0.5$	0.5	0.2	0.5	0.7	0.0175	✓	claim
$\mathbb{P}(Y=1)=0.5$	0.7	0.2	0.2	0.6	0.0084	✓	no_claim
$\mathbb{P}(Y=0)=0.5$	0.5	0.2	0.1	0.3	0.0015		
$\mathbb{P}(Y=1)=0.5$	0.7	0.2	0.2	0.6	0.0084	✓	
$\mathbb{P}(Y=0)=0.5$	0.5	0.2	0.5	0.3	0.0075		claim

. [2 marks]

From the table, the confusion matrix is as follows . . . . [0.5 mark]

	claim (0)	no_claim (1)
predict 0	1	2
predict 1	1	6

Accuracy : 0.7

Sensitivity : 0.5

Specificity : 0.75

Pos Pred Value : 0.3333

Neg Pred Value : 0.8571 . . . . [0.5 mark]

□

(c) (Ref: Tut 4 on Logistic Regression) Can we compare the logistic regression model in part (a) to the Naive Bayes model in part (b)? Can we say that the logistic regression model is better than the Naive Bayes model solely based on the performance metrics in part (a) and part (b)? Justify your answers with appropriate theory. (2 marks)

Solution. The two models cannot be compared because they are not trained with the same set of training data. . . . . [0.5 mark]

We cannot say that the logistic regression model is better because the testing data size is too small! . . . . [0.5 mark]

Theoretically, logistic regression model performs better with large number of data and the data is "linear". However, when the number of data is limited, Naive Bayes model will perform better than the logistic regression model based on Bayesian reasoning. [0.5 mark]

We need cross-validation in order to have a better understanding of the generalisation error. A single performance metric does not provide a good estimate for the generalisation error. [0.5 mark]

□