INF264 - Homework 8

Pierre Gillot **Natacha Galmiche***

Week 43 - 2021

Naive Bayes classifiers

In this exercise, we introduce the family of naive Bayes classifiers and get familiar with categorical naive Bayes with a simple toy dataset.

Naive Bayes classifiers are examples of Bayesian networks, that is probabilistic graphical models with a directed acyclic graph structure that encodes conditional independences among a set of random variables.

In naive Bayes classifiers, one typically represents the class C and a set of features $X_0, \dots, X_{d-1}$ as random variables and make the strong assumption that all features are conditionally independent from each other given the class:

$$\forall i \neq j : P(X_i, X_j|C) = P(X_i|C)P(X_j|C). \tag{1}$$

We can therefore represent the probability distribution of the class given the features in a compact form:

$$\begin{aligned} P(C|X_0, \dots, X_{d-1}) &= \frac{P(C, X_0, \dots, X_{d-1})}{P(X_0, \dots, X_{d-1})} \\ &= \frac{P(C)P(X_0, \dots, X_{d-1}|C)}{P(X_0, \dots, X_{d-1})} \\ &= \frac{P(C)\prod_{i=0}^{d-1} P(X_i|C)}{P(X_0, \dots, X_{d-1})}. \end{aligned} \tag{2}$$

Given a new sample $(x_0, \dots, x_{d-1})$ that we wish to predict, naive Bayes classi-

*Not a TA for this course this year.

fiers simply return the class maximizing $c \mapsto P(C\!=\!c|X_0\!=\!x_0,\dots,X_{d-1}\!=\!x_{d-1})$:

$$\begin{aligned} c_{\text{opt}} &= \operatorname*{argmax}_c P(C\!=\!c|X_0\!=\!x_0,\dots,X_{d-1}\!=\!x_{d-1}) \\ &= \operatorname*{argmax}_C \frac{P(C\!=\!c)\prod_{i=0}^{d-1} P(X_i\!=\!x_i|C\!=\!c)}{P(X_0\!=\!x_0,\dots,X_{d-1}\!=\!x_{d-1})} \\ &= \operatorname*{argmax}_C P(C\!=\!c)\prod_{i=0}^{d-1} P(X_i\!=\!x_i|C\!=\!c). \end{aligned} \tag{3}$$

1. Assuming all variables (class and features) are binary, how many probabilities do you need to compute in order to fully describe $P(C, X_0, \dots, X_{d-1})$ without the naive Bayes assumption ?

2. Same question under the naive Bayes assumption. Write the answer in the form $\mathcal{O}\big(f(d)\big)$, where you have to specify what f is.

3. Comment on the usefulness of naive Bayes.

In practice, what characterizes a naive Bayes classifier is:

- The specification of the set of variables (class and features).
- The priors on the probabilities $P(C)$ and $P(X_i|C)$.

Categorical naive Bayes classifiers

In categorical naive Bayes, we are given categorical features and class such that the probabilities $P(X_j\!=\!i|C\!=\!k)$ and $P(C\!=\!k)$ are estimated from the training set as

$$\begin{aligned} P(X_j\!=\!i|C\!=\!k) &\simeq \frac{N_{ijk}}{\sum_i N_{ijk}} \\ P(C\!=\!k) &\simeq \frac{N_k}{\sum_k N_k}, \end{aligned} \tag{4}$$

where N_{ijk} is the number of training examples with class k and jth feature taking value i, and N_k is the number of training examples with class k. This corresponds to the so-called maximum-likelihood estimator.

Suppose now that we have access to the following training dataset:

Weather	Temperature	Humidity	Wind	PlayTennis
Rainy	Hot	High	No	Yes
Rainy	Hot	High	Yes	No
Overcast	Hot	High	No	Yes
Sunny	Mild	High	No	Yes
Sunny	Cool	Normal	No	No
Sunny	Cool	Normal	Yes	No
Overcast	Cool	Normal	Yes	Yes
Rainy	Mild	High	No	No
Rainy	Cool	Normal	No	Yes
Sunny	Mild	Normal	No	Yes
Rainy	Mild	Normal	Yes	Yes
Overcast	Mild	High	Yes	Yes
Overcast	Hot	Normal	No	Yes
Sunny	Mild	High	Yes	No

We want to learn a model that is able to predict whether we should play tennis depending on the weather forecast. We were told we could easily train a categorical naive Bayes classifier for this problem.

4. What would be the variables of this categorical naive Bayes model ? Give their respective support.

5. Do you think the naive Bayes assumption is realistic for this particular model ?

6. You are given tomorrow's forecast:

   (Weather=Sunny, Temperature=Hot, Humidity=Normal, Wind=No)

   Would you go playing tennis tomorrow ?

7. The forecast for the day after tomorrow is still uncertain:

   (Weather=Rainy, Temperature=Hot, Humidity=?, Wind=?)

   Would you go playing tennis that day ?

8. Compute the naive Bayes estimate

   P(PlayTennis=No|Weather=Overcast, Temperature=Cool, Humidity=High, Wind=Yes).

   Is this estimate satisfying ? What caused it ?

9. Assume now that we slightly "smooth" the estimator of the conditional probabilities $P(X_j = i | C = k)$:

$$P(X_j = i | C = k) \simeq \frac{N_{ijk} + \alpha_{ijk}}{\sum_i (N_{ijk} + \alpha_{ijk})}. \tag{5}$$

Recompute the naive Bayes estimate

P(PlayTennis=No|Weather=Overcast, Temperature=Cool, Humidity=High, Wind=Yes)

assuming $\forall i, j, k : \alpha_{ijk} = 1$.

10. What is the role of α ? When is it particularly important to tune this hyperparameter ?

11. Verify your results in sklearn with the class 'sklearn.naive_bayes.CategoricalNB'.