CZ4041/CE4041: Machine Learning

Lesson 2b: Bayesian Decision Theory

Sinno Jialin PAN

School of Computer Science and Engineering, NTU, Singapore

Decisions with Posteriors: Limitation

- By far, a decision (or prediction) is made based on the maximum posterior
  - Cost of misclassification on different classes is not taken into consideration
- However, in some application domains, like the medical domain, the cost of misclassification on different classes may be different

An Example

- To diagnose whether a patient A is with covid-19: $y = 1$ (Yes) or $y = 0$ (No). Suppose based on a trained Bayesian classifier, we know that $P(y = 1|x_A) = 0.1$. Should the doctor diagnose that A is with covid-19 or not?
  - Cost of misclassifying a healthy patient with covid-19:

Stay in hospital
and take more tests

  - Cost of misclassifying a patient with covid-19 as healthy:

Community outbreak!

Loss or Cost

- Actions: a_c, i.e., predict $y = c$, where $c = 0, \ldots, C-1$
- Define λ_{ij} as the loss/cost of a_i when the optimal action is a_j (i.e., predict $y = i$ while true class label is j)
- E.g., in the previous example, $y = 0$: healthy, and $y = 1$: with covid-19 (binary classification)
- We define two corresponding actions: a_0: predict $y = 0$ and a_1: predict $y = 1$, and the losses as

$$\begin{cases} \lambda_{00} = 0 & \text{predict correctly} \\ \lambda_{11} = 0 & \text{predict correctly} \\ \lambda_{01} = 10 & \text{misclassify 1 as 0} \quad \text{(misclassify with covid-19 as healthy)} \\ \lambda_{10} = 1 & \text{misclassify 0 as 1} \quad \text{(misclassify healthy as with covid-19)} \end{cases}$$

Expected Risk

- Expected risk for taking action a_i:

$$R(a_i|\boldsymbol{x}) = \sum_{c=0}^{C-1} \lambda_{ic} P(y = c|\boldsymbol{x})$$

- Explanation: to estimate a risk of taking an action, one needs to consider all the possible losses
  - Specifically, taking action a_i (predict $\boldsymbol{x}$ belonging class i), as the ground-truth label of $\boldsymbol{x}$ can be any class in the C classes, we need consider all the possible losses: $\lambda_{i0}, \lambda_{i1}, \dots, \lambda_{i(C-1)}$
  - We can simply use the average $\frac{\lambda_{i0}+\cdots+\lambda_{i(C-1)}}{C}$ to estimate its risk
  - The possibilities of each loss occurring are different because the probabilities that $\boldsymbol{x}$ belongs to each class are different
    - Use the $P(y = c|\boldsymbol{x})$ as a weight for each loss λ_{ic}, and compute the weighted sum of all possible losses → expected risk

An Example

- Expected risk for taking action a_i: $R(a_i|\boldsymbol{x}) = \sum_{c=0}^{C-1} \lambda_{ic} P(y = c|\boldsymbol{x})$
- Consider the covid-19 example
  - $P(y = 1|\boldsymbol{x}_A) = 0.1$ and $P(y = 0|x_A) = 0.9$
  - a_0: predict $y = 0$ (healthy), and a_1: predict $y = 1$ (with covid-19)

$$\begin{cases} \lambda_{00} = 0 & \text{predict correctly} \\ \lambda_{11} = 0 & \text{predict correctly} \\ \lambda_{01} = 10 & \text{misclassify 1 as 0} \quad \text{(misclassify with covid-19 as healthy)} \\ \lambda_{10} = 1 & \text{misclassify 0 as 1} \quad \text{(misclassify healthy as with covid-19)} \end{cases}$$

  - Expected risk of taking action a_0 (predict patient A as healthy)

    $R(a_0|\boldsymbol{x}_A) = \lambda_{00} P(y = 0|\boldsymbol{x}_A) + \lambda_{01} P(y = 1|\boldsymbol{x}_A) = 1$
  - Expected risk of taking action a_1

    $R(a_1|\boldsymbol{x}_A) = \lambda_{10} P(y = 0|\boldsymbol{x}_A) + \lambda_{11} P(y = 1|\boldsymbol{x}_A) = 0.9$

Decision based on Expected Risk

- Choose the action with minimum risk:

$$\text{Choose } a^* \text{ if } a^* = \arg\min_{a_c} R(a_c|\boldsymbol{x})$$

- In the covid-19 example
  - Expected risk of taking action a_0 (predict as healthy)

$$R(a_0|\boldsymbol{x}_A) = 1$$

  - Expected risk of taking action a_1 (predict with covid-19)

$$R(a_1|\boldsymbol{x}_A) = 0.9$$

  - Thus, we choose action a_1: predict patient A is more likely with covid-19

A Special Case

- Making predictions based on maximum posterior is a special case of making decisions based on minimum expected risk
- Define the losses as

$$\lambda_{ij} = \begin{cases} 0 & \text{if } i = j \\ 1 & \text{if } i \neq j \end{cases}$$

All correct decisions have no loss and all errors are equally costly

Known as the 0/1 loss

A Special Case (cont.)

- With the 0/1 loss: $\lambda_{ij} = \begin{cases} 0 & \text{if } i = j \\ 1 & \text{if } i \neq j \end{cases}$
- The expected risk of taking action a_i:

$$R(a_i|\boldsymbol{x}) = \sum_{c=0}^{C-1} \lambda_{ic} P(y = c|\boldsymbol{x})$$

$$= \lambda_{i0} P(y = 0|\boldsymbol{x}) + \cdots + \overset{=0}{\boxed{\lambda_{ii}}} P(y = i|\boldsymbol{x}) + \cdots + \lambda_{i(C-1)} P(y = C-1|\boldsymbol{x})$$

$$= P(y = 0|\boldsymbol{x}) + \cdots + P(y = i-1|\boldsymbol{x}) + P(y = i+1|\boldsymbol{x}) \ldots + P(y = C-1|\boldsymbol{x})$$

$$= \sum_{j \neq i} P(y = j|\boldsymbol{x}) = \boxed{\sum_{c} P(y = c|\boldsymbol{x})} - P(y = i|\boldsymbol{x}) = 1 - P(y = i|\boldsymbol{x})$$

$$\sum_{c} P(y = c|\boldsymbol{x}) = 1$$

A Special Case (cont.)

- The expected risk of taking action a_i:

$$R(a_i|\boldsymbol{x}) = 1 - P(y = i|\boldsymbol{x})$$

- Choose an action with minimum expected risk,

$$\text{Choose } a_i \text{ if } R(a_i|\boldsymbol{x}) = \min_{a_c} R(a_c|\boldsymbol{x})$$

$$\Updownarrow \text{ Equivalent to}$$

$$\text{Predict } y = c^* \text{ if } P(y = c^*|\boldsymbol{x}) = \max_{c} P(y = c|\boldsymbol{x})$$

Bayesian Decision Theory: Summary

- If cost of misclassification on different classes is available, rather than only using posterior probabilities (usually estimated by a Bayesian classifier), Bayesian decision theory provides a way to encode the cost information into decision making

We will introduce the first example of Bayesian Classifier: Naïve Bayes Classifiers next week

Thank you!