

**3.5 )** *Propose a three-level cascade where when one level rejects, the next one is used as in Equation 3.10. How can we fix the  $\lambda$  on different levels?*

For complex decisions, such as misclassifications, questionable outliers, or missing portions of data, a cascade or piece-wise approach can be used classify the data. Using multiple classes, an algorithm can utilize an additional action, compared to a binary classification, to reject or doubt when deciding the data classification.

Let's assume we have  $K$  number of classes and  $i$  represents the  $i^{\text{th}}$  class of  $C$ . The expected risk of taking an action  $\alpha_i$  during the classification process for data  $\mathbf{x}$  can be obtained by the following equation:

$$R(\alpha_i|\mathbf{x}) = \sum_{k=1}^K \lambda_{ik} P(C_k|\mathbf{x})$$

where,  $\lambda_{ik}$  represents the loss incurred by performing action  $\alpha_i$ . We wish to implement a three-level classification scheme where the level either accepts the data or rejects the data to the next class. This is can be thought as analogous to an *else if*, or *elif* in Python, statement in computer programming, where each *elif*, or classification level in our scenario, introduces a new statement that if true performs action and ends the loop.

$$\lambda_{ik} = \begin{cases} 0 & \text{if } (i = k) \\ \lambda & \text{if } (i \neq k \text{ and } i = K + 1) \\ 1 & \text{if } (i \neq k \text{ and } i \neq K + 1) \end{cases}$$

Similar to that of Eq. 3.10 in textbook<sup>1</sup> where  $0 < \lambda < 1$  represents the loss from  $\alpha_{K+1}$  action. This proposed method provides an example three-level cascade, where one level rejects and the following level is used.

---

<sup>1</sup>Alpaydin, E. *Introduction to Machine Learning*. 3<sup>rd</sup> Edition. MIT Press, Cambridge, MA. 2014. ISBN-978-81-5078-6.

**3.9 )** Show that as we move an item from the antecedent to the consequent, confidence can never increase:  $\text{confidence}(ABC \rightarrow D) \geq \text{confidence}(AB \rightarrow CD)$ .

An association rule is an implication of the form  $X \rightarrow Y$  where  $X$  is the antecedent and  $Y$  is the consequent of the rule. There are multiple measures of calculated for learning associations, but conventional two measures are primarily used: *support* and *confidence*. Generally speaking, the support measure shows the statistical significance of the rule and confidence shows the strength of the rule.

The confidence of the association rule  $X \rightarrow Y$ :

$$c(X \rightarrow Y) \equiv P(Y|X) = \frac{P(X, Y)}{P(X)} \text{ where, } c \text{ represents the confidence.}$$

Confidence of rules generated from the same item set has an antimonotone property, such as For our problem, probability functions are:

$$c(ABC \rightarrow D) \equiv P(D|ABC) = \frac{P(ABC, D)}{P(ABC)}$$

$$c(AB \rightarrow CD) \equiv P(CD|AB) = \frac{P(AB, CD)}{P(AB)}$$

For the sake of clarity, it is beneficial to expand the confidence equations into the probability component. Let's assume we are using basket analysis, where a we have four items ( $A, B, C, D$ ) and a database of 500 purchases. In this scenario, the confidence is the conditional probability that a randomly selected transaction will include all the items in the consequent, given that the transaction includes all the items in the antecedent.

$$c(ABC \rightarrow D) = \frac{P(ABC, D)}{P(ABC)} = \frac{\#\{\text{customers who bought ABC and D}\}}{\#\{\text{customers who bought ABC}\}} = \frac{\#\{\text{customers who bought ABCD}\}}{\#\{\text{customers who bought ABC}\}}$$

$$c(AB \rightarrow CD) = \frac{P(AB, CD)}{P(AB)} = \frac{\#\{\text{customers who bought AB and CD}\}}{\#\{\text{customers who bought AB}\}} = \frac{\#\{\text{customers who bought ABCD}\}}{\#\{\text{customers who bought AB}\}}$$

Table 1: Single Item purchases.

Item	Purchases
A	150
B	200
C	50
D	100

Table 2: Double item purchases.

Item	Purchases
AB	50
AC	5
AD	15
BC	10
BD	20
CD	20

Table 3: Third item purchases.

Item	Purchases
ABC	5
ABD	3
ACD	1
BCD	2

Table 4: Four item purchases.

Item	Purchases
ABCD	1

$$c(ABC \rightarrow D) = \frac{\#\{\text{customers who bought ABC and D}\}}{\#\{\text{customers who bought ABC}\}} = \frac{1}{5} = 0.2$$

$$c(AB \rightarrow CD) = \frac{\#\{\text{customers who bought AB and CD}\}}{\#\{\text{customers who bought CD}\}} = \frac{1}{20} = 0.05$$

This example proves that as we move an item from the antecedent to the consequent the confidence cannot improve. Moreover, as we continue to decrease the sample of the population by increasing the group size (e.g.  $P(A|B)$  to  $P(AB|C)$ ) for our data, each data point holds more weight, thus larger sample contribution.

**3.10 )** *Associated with each item sold in basket analysis, if we also have a number indicating how much the customer enjoyed the product, for example, on a scale of 0 to 10, how can you use this extra information to calculate which item to propose to a customer?*

Using the scale value (0 – 10) for each customer for a particular variable, we can begin to build a correlation matrix for the variables, which can be used to study the correlations between products and a particle variable, such as amount of enjoyment. Using this correlation, the store can begin to recommend similar items or items that are typically correlated with the original purchased item that provided the customer enjoyment. A decision threshold value can be utilized to recommend items, such that a low level of enjoyment decreases the probability of recommending similar products and vice versa, such as a value of 7.0 out of 10.0.