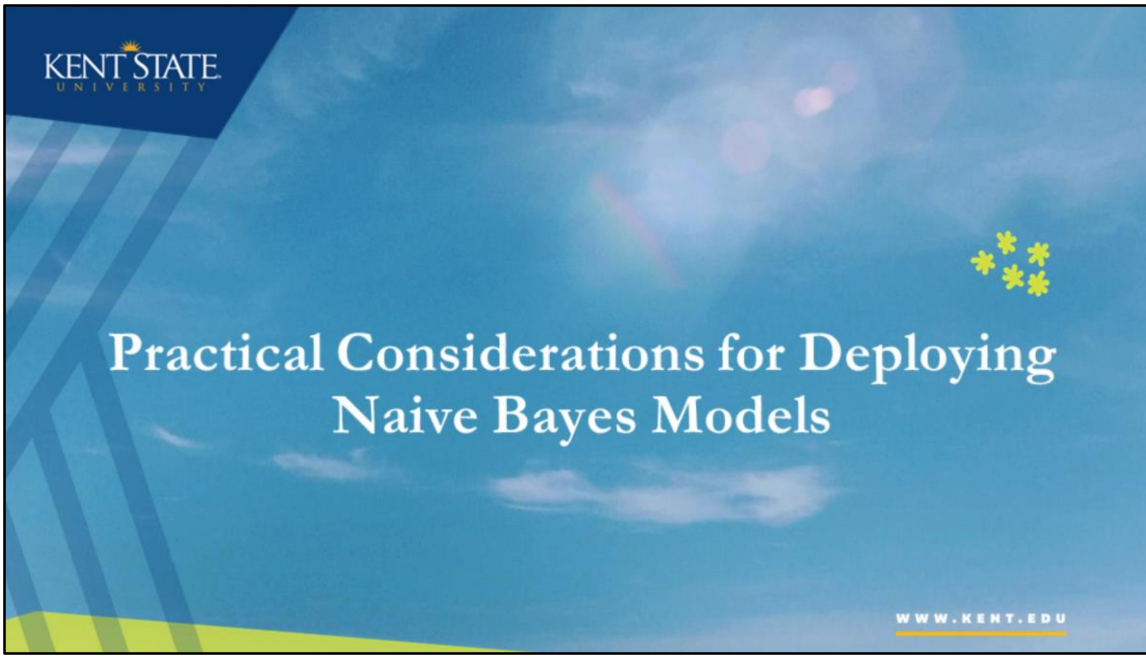


Let us now look at some practical considerations in applying NB

KENT STATE UNIVERSITY

Assumption of Normality

- Recall that Naïve Bayes model assumes that all features are independent given the class label Y. That is to say

$$P(X_1, \ldots, X_n|Y) = \prod_{i=1}^{n} P(X_i|Y)$$

- For categorical variables, this calculation is very straightforward.
- However, when dealing with numerical variables, it is often assumed that data follows a Gaussian distribution, so that to make it easy to calculate probabilities.
- This assumption, however, may not hold true all the time.

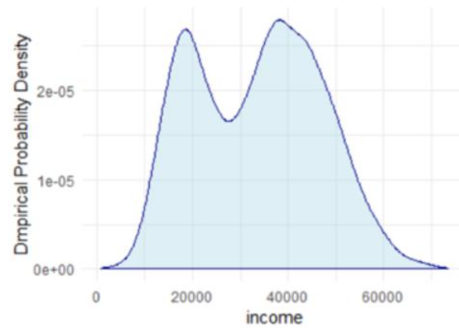
WWW.KENT.EDU

Generally, the NB model is only applied when all predictor variables are categorical. This allows us to calculate the conditional probabilities easily. We could also calculate the conditional probabilities for numeric predictors if we make some assumptions on their distributional forms. One such form is to assume that the distribution is Gaussian (Normal).

KENT STATE UNIVERSITY

Example

- Let us consider again the credit card default example.
- The distribution of income does not follow a bell shape Gaussian distribution.
- Instead, we can see a bi-modal distribution with two peaks.



WWW.KENT.EDU

For this example, it is clear that the distribution of data is not normal. We could transform the data to make it Normal. This is similar to how in regression we transform variables to satisfy the normality assumption.

Box-Cox Transformation

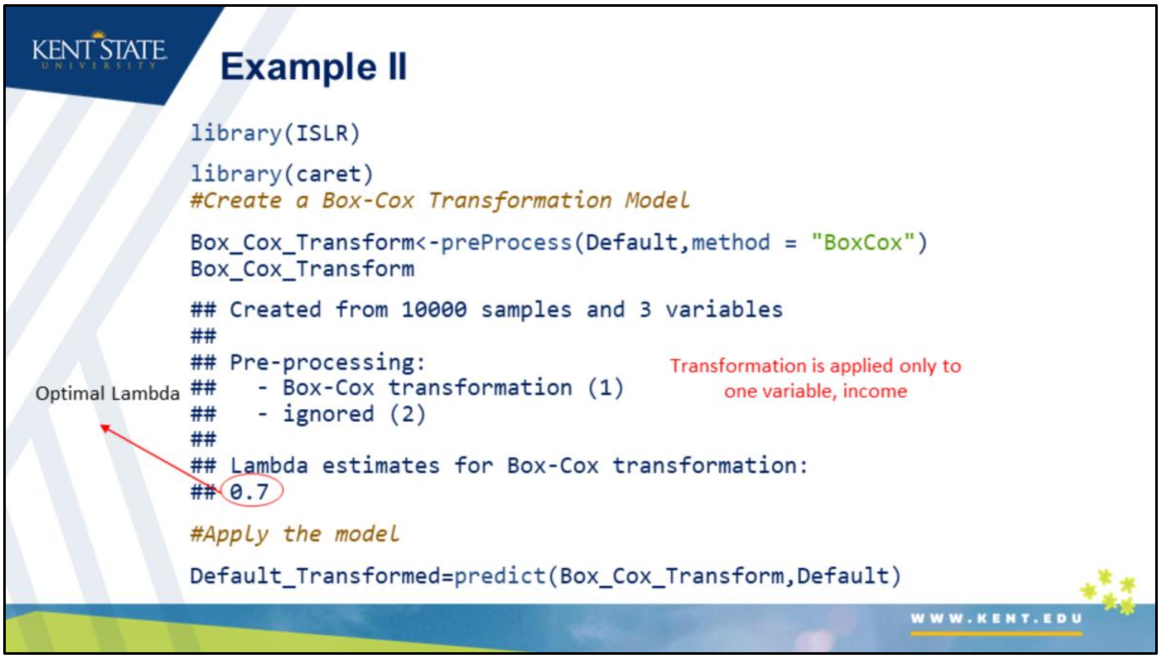
- Box-Cox transformation can be applied to make a non-normal distribution to look more like a normal distribution.
- The Box-Cox transformation of the variable x is also indexed by λ, and is defined as

$$x'_\lambda = \frac{x^\lambda - 1}{\lambda}.$$

- The preProcess() function in 'caret' package can be used to apply the Box-Cox transformation by finding the optimal value of λ

WWW.KENT.EDU

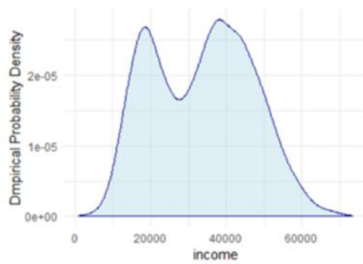
The box-cox approach allows us to determine the transformation that is needed to get non-normal data to look more normal. This transformation is part of the "caret" package.



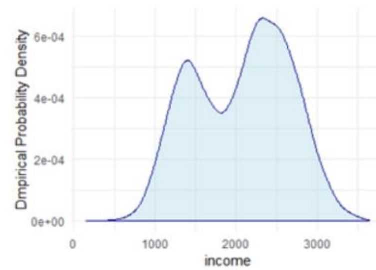
The Box-Cox transformation here suggests a value of 0.7 for lambda, where the transformation is applied to the income variable. In other words, we transform income to (income^.7 - 1) / 0.7.

Box-Cox Transformation Distribution

- Box-Cox transformation, however, may not be sufficient to make a variable normally distributed.



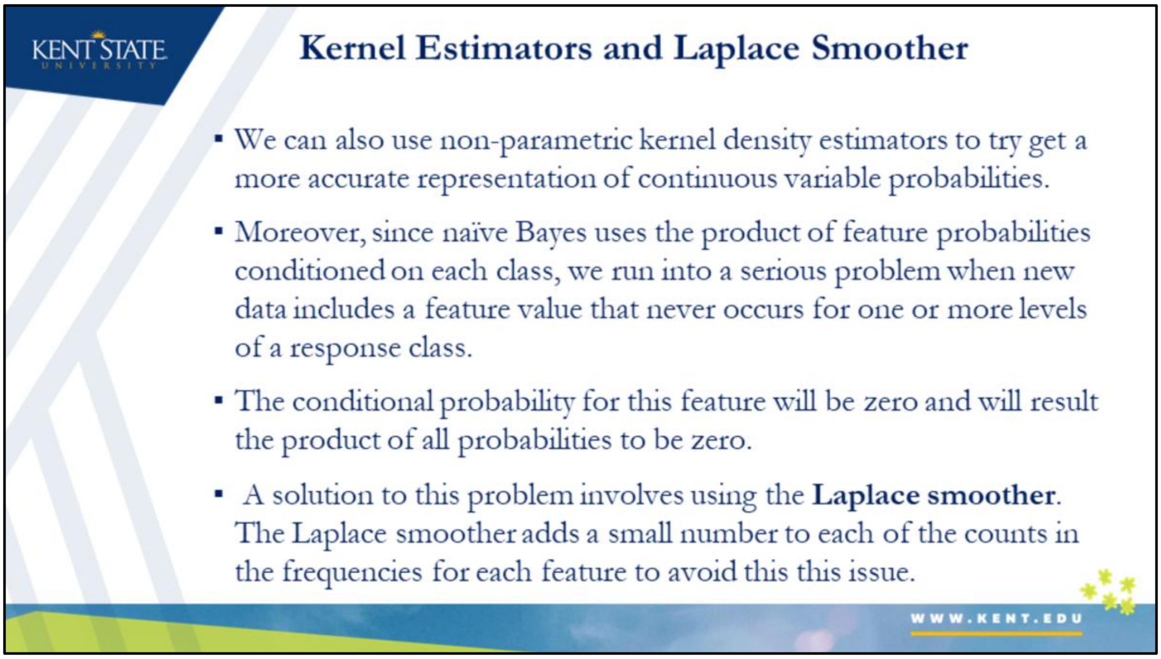
Income variable, original



Income variable, box-cox transformed. Slightly improved but still far from a Normal distribution

WWW.KENT.EDU

Notice that while the distribution looks more normal, it is still a bi-modal distribution. In such case, stronger measures are needed to satisfy assumptions.

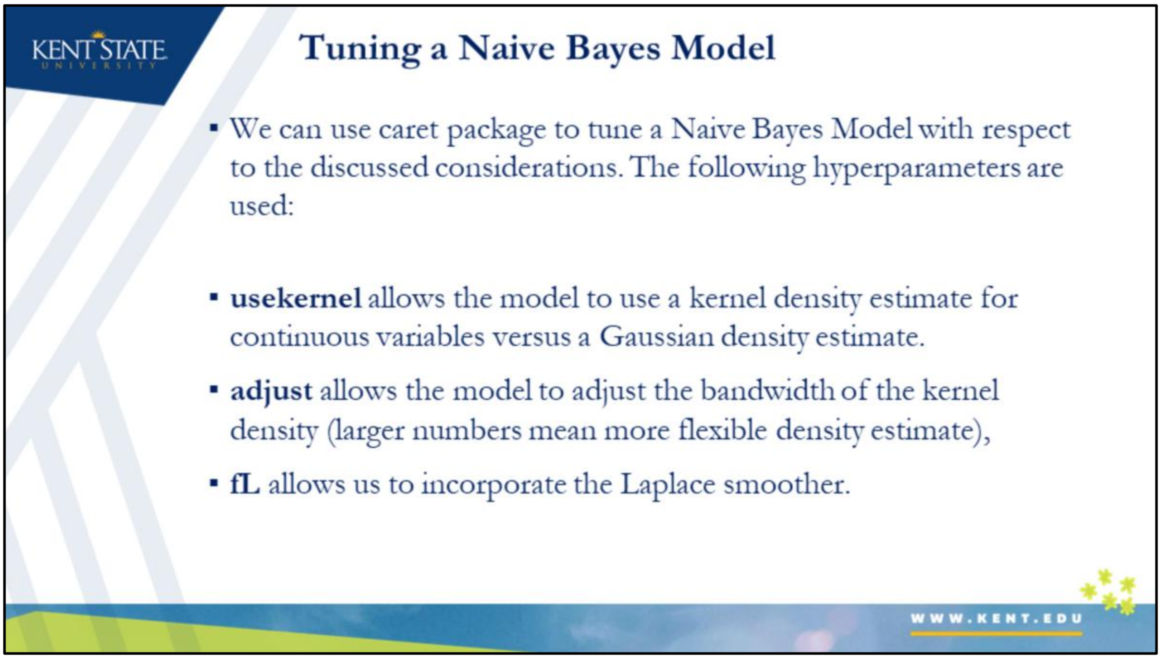


There are three main difficulties in applying the NB model.

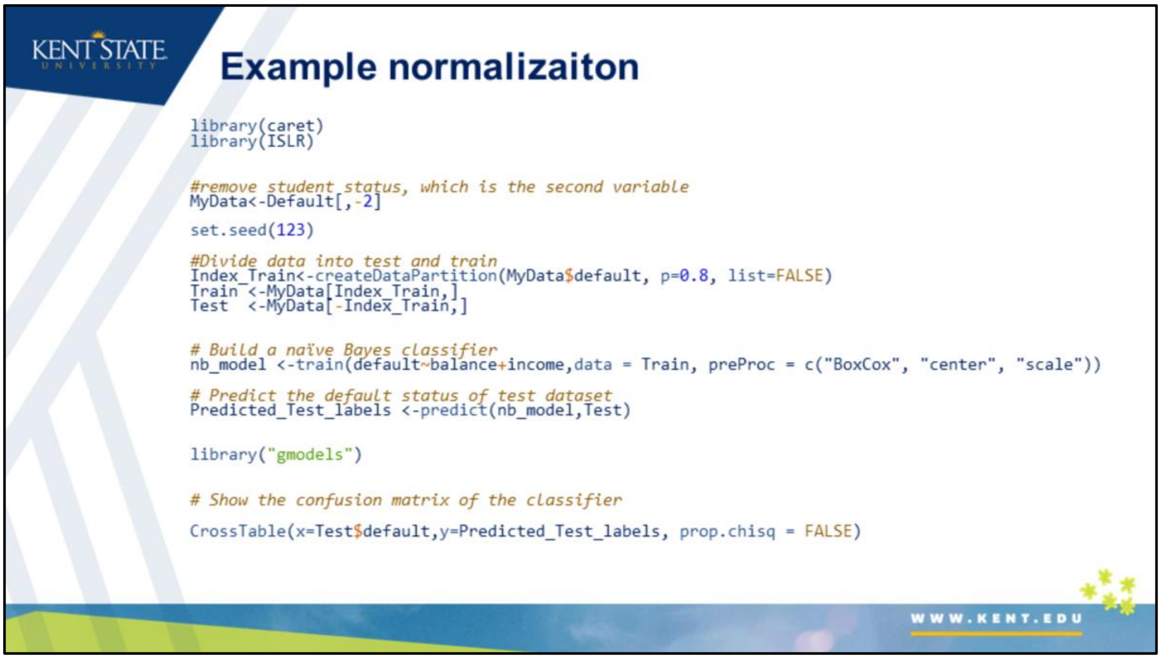
First, the naive Bayes classifier requires a very large number of records to obtain good results.

Second, where a predictor category is not present in the training data, naive Bayes assumes that a new record with that category of the predictor has zero probability. This can be a problem if this rare predictor value is important. A popular solution in such cases is to replace zero probabilities with non-zero values using a method called smoothing (e.g., Laplace smoothing can be applied by using argument laplace = 0 in function naiveBayes()).

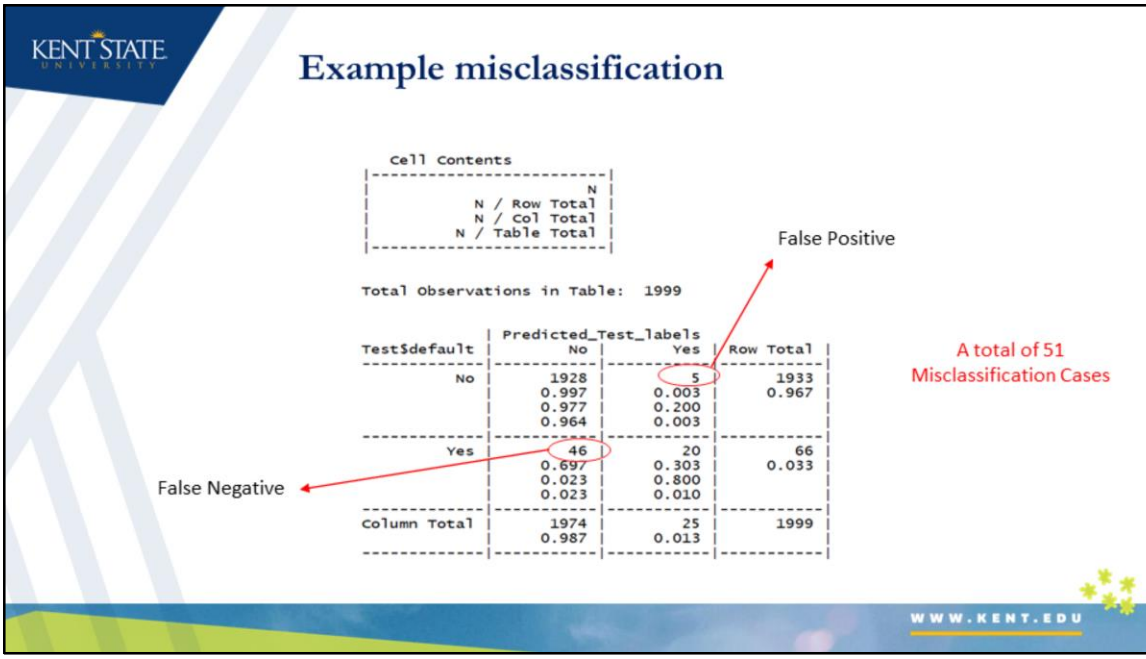
Finally, good performance is obtained when the goal is classification or ranking of records according to their probability of belonging to a certain class. How- ever, when the goal is to estimate the probability of class membership (propensity), this method provides very biased results. For this reason, the naive Bayes method is rarely used in credit scoring (Larsen, 2005).



Hypertuning is always an important aspect of improving model performance. The NB model is no different. This slide provides some hypertuning parameters.



Here, we both normalize the data, and also use the Box-Cox transformation.



Example misclassification

```
   Cell Contents
|-------------------------|
|                       N |
|           N / Row Total |
|           N / Col Total |
|         N / Table Total |
|-------------------------|

Total Observations in Table:  1999

             | Predicted_Test_labels
 Test$default |        No |       Yes | Row Total |
-------------|-----------|-----------|-----------|
          No |      1928 |         5 |      1933 |
             |     0.997 |     0.003 |     0.967 |
             |     0.977 |     0.200 |           |
             |     0.964 |     0.003 |           |
-------------|-----------|-----------|-----------|
         Yes |        46 |        20 |        66 |
             |     0.697 |     0.303 |     0.033 |
             |     0.023 |     0.800 |           |
             |     0.023 |     0.010 |           |
-------------|-----------|-----------|-----------|
 Column Total |      1974 |        25 |      1999 |
             |     0.987 |     0.013 |           |
-------------|-----------|-----------|-----------|
```


This produces a slight improvement in misclassification rates.

This concludes our presentation for the NB model.