**How does your error rate here compare with the error rate you would’ve obtained if you had used a k-nearest neighbors’ model to predict outcomes on test? Aside from error rate, why might we prefer Naıve Bayes’ over kNN, or vice versa?**

|  |  |  |
| --- | --- | --- |
|  | **kNN** | **Naïve Bayes** |
| **error\_rate** | 8.90% | 10.1% |
| **False Positives** | 44 | 2 |
| **False Negatives** | 134 | 200 |
| **sensitivity rate** | 25.40% | 0.5% |
| **specificity rate** | 98.30% | 99.8% |
| **benchmark error** | 9.95% | 10.05% |

The error of knn is almost 1% more than that model’s benchmark error. The error of Naïve Bayes model is 0.05% more than that model’s benchmark error. I would say that means that Naïve Bayes performs slightly better given the context.

We would prefer kNN when we have numerical inputs and we want to predict a categorical output. (Note: It’s not like kNN should only be used if input is numeric. We can even change numeric data to factors by grouping them into bins).

kNN regression helps output data points identify the closest training observation point. Prediction for a point is entirely dependent on its neighbors. Any application where movements are synchronous, like stock movements in NASDAQ or S&P500 move in sync with market shocks, we can use kNN to predict the movement or value of the target. In the case of Universal Bank data, we wanted to predict the target (whether customer will borrow) based on experience, avg spending on credit card, age, mortgage and income. With knn analysis, the output doesn’t show us which attribute has the greatest/least affect. kNN is non-parametric and hence not interpretable – so you cannot answer questions like why is this point classified as Borrow instead of Not Borrow? OR what is the relationship between Personal loan and Online.

Naïve Bayes would be used when we have categorical inputs and we want to predict categorical output. Here, we are trying to find the probability of target output i.e. customers borrowing, given certain conditional inputs – i.e. whether he opened a securities account, whether he has a credit card, whether he has a CD account and/or whether he engages in online banking. We have a better understanding of how other factors influence the probability of people borrowing and how borrowing influences the probability of people engaging in other activity – i.e. it is a two way street.