



Investigating the Performance of Naive- Bayes Classifiers and K- Nearest Neighbor Classifiers

Mohammed Jahirul Islam
Dept. of Elec. & Comp. Engineering
University of Windsor, Windsor, ON, Canada



Presentation Outline

- **Overview of Classification**
- **Problem Statement and Motivation**
- **Literature Review**
 - Bayes Classifier
 - Naïve Bayes Classifier
 - K- Nearest Neighbor Classifier
- **Application of Classifiers**
 - **Credit Card Approval**
 - Experimental Results
- **Conclusion, Comments**



Classification- Overview

- Goal of machine learning is to program computers to use example data or past experience to solve a given problem
- Classification is an application of machine learning
- Takes raw data and classifies it as belonging to a particular class based on required parameter set
- **Selection of right classification algorithm for machine learning is a big issue**



Classification Scheme

- Selection of classifier depends on the application and the information available from that application
- Machine learning uses the theory of statistics in building mathematical model for classification, because the core task is making inference from a sample
 - Inference is a big deal



Problem Statement

- Key Question?
 - Is there any way to generalize the classification techniques?
 - How to determine which technique is suitable for a specific problem?
 - How to improve a specific classifier by changing the parameters for a specific application?
- **Investigating the performance of the classifiers could be one solution to reach that goal.**

Literature Review

- A wide range of algorithms are available for classification from Bayesian classifiers to more powerful Neural Network.
- Bayesian theory is basically works as a framework for making decision under uncertainty- a probabilistic approach to inference
- The probability of the future events could be calculated by determining the earlier frequency:
 - To see the future, look at the past
- The predictions are based completely on data culled from reality
 - The more data obtained, the better it works
- **Bayesian models are self-correcting**
 - **When data changed, so do the result**



Literature Review (cont'd)

- In classification, Bayes rule is used to calculate the probabilities of the classes and it is a big issue how to classify raw data rationally to minimize expected risk.
- What if the dimension of the inputs is so high?
- Naïve Bayes classifier is one of the mostly used practical Bayesian learning methods.
 - Very effective when the dimensionality of the input is very high
- In some domains, it's performance is comparable to that of neural network
- K- Nearest Neighbor algorithm is the most basic instance-based method
 - Store the training instances in look up table and interpolate from these

Bayesian Theory

- Most practical learning approach for most learning problems
 - Based on evaluating explicit probabilities for hypothesis
- Bayes theorem states that:

$$P(h | D) = \frac{P(D | h)P(h)}{P(D)}$$

- $P(h)$: Prior probability of hypothesis h - prior
 - $P(D)$: Prior probability of training data D - Evidence
 - $P(D|h)$: Probability of D given h - Likelihood
 - $P(h|D)$: Probability of h given D - Posterior probability
-
- The posterior probability of class h_i is calculated and finally the best hypothesis (h_{MAP}) is selected- Maximum a posteriori probability

Naïve Bayes Classifier

- It requires a small amount of training data to estimate the parameters necessary for classification
- Highly practical Bayesian learning method
 - Particularly suited when dimensionality of the input is so high
- Assumption:
 - The attribute values are conditionally independent given target value
 - It ignores the possible dependencies, say correlations among input
- Reduce a multivariate problem to a group of univariate problems



K- Nearest Neighbor Classifier

- In parametric methods, we assume a model is valid for over the whole input space
 - Practically this assumption does not hold and we may incur a large error if it does not, solution?
- In nonparametric estimation we assume similar inputs have similar outputs.
 - It does not use any model to fit data
 - Based on memory/ training data.
 - Called instance-based/ memory based learning algorithm
- **KNN is instance-based classifier**

KNN Classifier

- KNN is the most basic instance-based learning method
 - Result of new instance query is classified based on majority of KNN category
 - Assumption: The world is so smooth and functions changes slowly.
- Find the similar past instances from the training set
 - Uses suitable distance measure, k
- It is common to select k small and odd to break ties (typical value 1, 3, 5)
 - Larger k values help reduce the noisy points

Implementation, Training and Testing

- Naïve Bayes and KNN classifiers are implemented to apply “Credit Card Approval” application.
- It is important for a bank/ financial institution to be able to predict in advance the risk associated with a loan
 - The probability that the customer will default and not the whole amount back
- Make sure that the bank will make profit and also to not inconvenience a customer with his/her financial capacity.
- Usually, the information about the customer includes income, savings, collaterals, profession, age, passed financial history and so forth

Story Behind the Datafile

- The source of the datafile:
 - *ftp://ftp.ics.uci.edu/pub/machine-learning-databases/credit-screening*
- All attribute names and values have been changed to meaningless symbols to protect confidentiality of the data.
- Contains information about 671 applicants, whether they were approved or rejected.
- Each application is described by 9 attributes and classified as approved (“+”) or rejected (“-”)
- Each of the 9 attributes is a one letter symbol that is a shorthand for a more meaningful English language description.



Attributes of the Datafile

Attributes	Value
A1	b, a
A2	u, y
A3	g, p, h
A4	i, k, c, g, q, d, a, m, x, w, j, r, e, b
A5	h, v, f, d, b, j, z, m, o
A6	t, f
A7	t, f
A8	t, f
A9	g, p, s

Experimental Results

- Training set examples: 470, Testing set example: 201
- Using Naïve Bayes classifier at first the tables are constructed from the attributes A1 to A9 using the training set
- Sample table is shown for attribute A9

Table: $P(A9|accept)$ and $P(A9|reject)$, Total accept: 215, Reject: 255

A9	Accept	Reject
g	0.9581	0.9020
p	0.0047	0.0039
s	0.0372	0.0941

Experimental Results (Naïve)

- The test set is classified based on the probabilities estimated from training set
- Each example is picked from testing set and then its class is predicted
- The predicted class is compared to the target value that is given in the test set.
- If mismatch, example becomes error
- The classification is shown in Table, Total testing example: 201

Classification	Number	%
Correct	176	87.57
Incorrect	25	12.43

Results- KNN

- Different values of k are tried, for example $k=1, 3, 5, 11, 51, 101$
- Distance metric is calculated- Euclidian distance, for mismatch distance set to 1, else 0
- Based on the value of k , the k number of smallest distance training examples are picked up and calculate their corresponding accept or reject
- The larger value is the predicted value for this testing example and it is compared to the target value. If mismatch, example becomes error



Results- KNN (Cont'd)

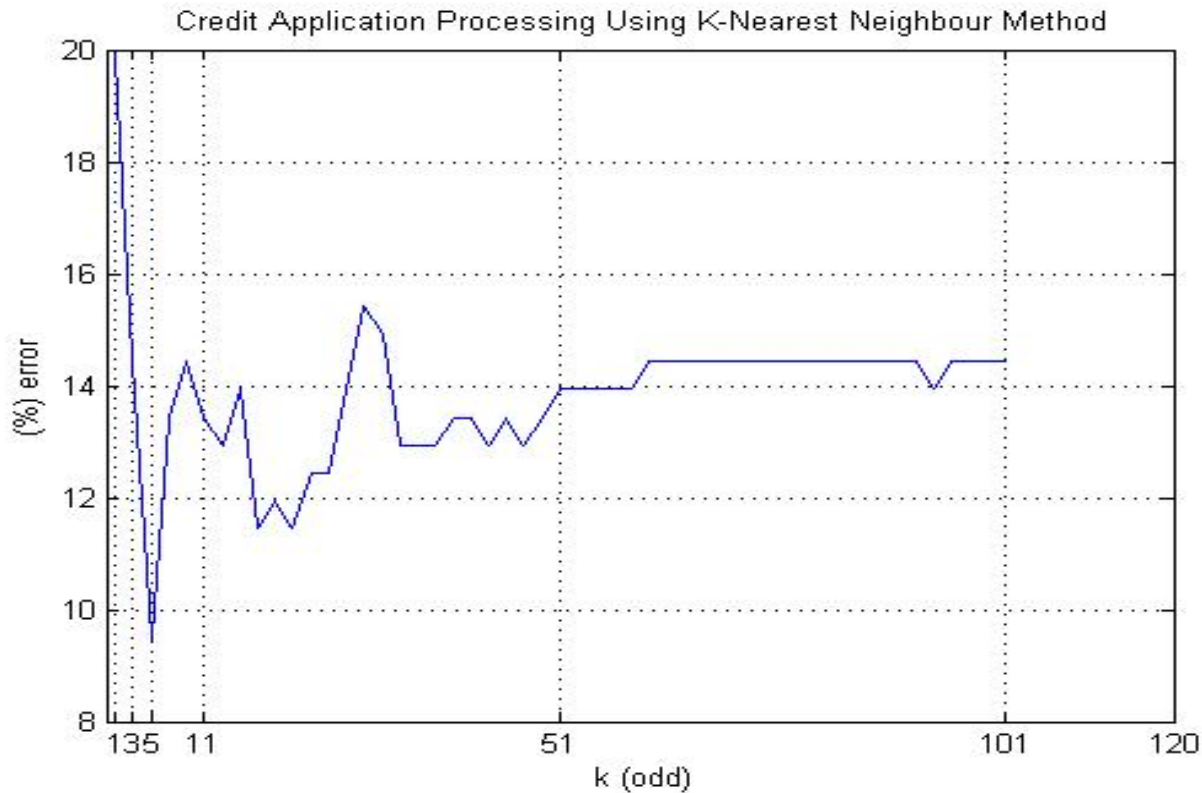
- For different values of k, the testing set is classified
 - % of error for different k using KNN, testing example: 201

K	Correct (%)	Incorrect (%)
1	80.10	19.90
3	85.57	14.43
5	90.55	9.45
11	86.57	13.43
51	86.07	13.93
101	85.57	14.43



Results- KNN (Cont'd)

- For different values of k , the testing set is classified
 - % of error for different k using KNN, testing example: 201



Comparative Statement (KNN and Naive)

- Naïve Bayes and KNN classifier is compared in terms of correct classification and misclassification rate
- The classification is shown in Table, Total testing example: 201
-

Classifier	Correct Classification		Misclassification	
	Number	%	Number	%
Naïve	176	87.57	25	12.43
KNN (k=5)	<u>182</u>	<u>90.55</u>	<u>19</u>	<u>9.45</u>

Conclusions (Bayesian)

- 'Credit card approval' application is selected for investigating the performance of widely used classifiers Naïve Bayes and KNN classifier.
- The result of Bayesian inference depends strongly on prior probabilities.
- Bayes theorem provides a principled way to calculate the posterior probability of each hypothesis given the training data and select the most probable one

Conclusions (Naïve Bayes, KNN)

- Naïve Bayes classifier is applied to the credit card approval testing data set and found 12.43% error of classification
- Instance-based methods are sometimes referred to as “Lazy” learning methods, because they delay the process until a new instance must be classified
- In KNN, the selection of K is application dependent. To simplify the problem, it was fixed to odd number so that no tie can happen
- At $K=5$, the misclassification rate is 9.45% (minimum), so $k=5$ is the best value for the application



References

- [1] <ftp://ftp.ics.uci.edu/pub/machine-learning-databases/creditscreening>.
- [2] E. Alpaydin. “*An Introduction to Machine Learning.*”, The MIT press, Cambridge, Massachusetts, London, England, 2004.
- [3] T. Cover and P. Hart. “Nearest neighbor pattern classification”, *IEEE Transaction on Information Theory*, 13:21–27, 1967.
- [4] R. Duda, P. Hart, and D. Stork. “*Pattern Classification*”, Wiley Interscience, 2nd ed.
- [5] S. Eyheramendy, D. Lewis, and D. Madigan “On the naïve bayes model for text categorization”, *Proceedings Artificial Intelligence Statistics*, 2003.
- [6] D. Lewis “Naive (Bayes) at Forty: The Independence Assumption in Information Retrieval”, ATT lab Research, NJ, USA.
- [7] T. Mitchell. *Machine Learning*. McGraw-Hill.
- [8] I. Rish “An empirical study of the naive bayes classifier”, *Proceedings of IJCAI-01*, 2001.
- [9] N. Roussopoulos, S. Kelley, and F. Vincent “Nearest neighbor queries”, *Proceedings of the 1995 ACM SIGMOD International Conference on Management of Data*, 1995.
- [10] K. Weise and W. Woger “Comparison of two measurement results using the Bayesian theory of measurement uncertainty”, *Measurement Science and Technology*, 5:879–882, 1994.
- [11] Q. J. Wu. *Class Notes- Machine Learning and Computer Vision*. University of Windsor, Windsor, ON, Canada, 2007.



Thanks for your patience

Questions?