

CAP5638 Project 1

Classification Using Maximum-likelihood, Parzen Window, and k-Nearest Neighbor

Suhib Sam Kiswani

October 26, 2015

TODO

You need to turn in a report, including description of your algorithms and implementations, experimental results, and comparison and evaluation and analysis of the three different methods.

Abstract

For each dataset, you need to first estimate the necessary data normalization and parameters using the training set only (you need to document the results for different choices of parameters on the training set in the leaveone- out sense for Parzen windows and k-nearest neighbor rule) and then document the classification results on the test set using each of the methods. Then you need to compare different methods in terms of classification performance and required time for classification, and give justifications for your observed empirical results.

1 Implementation Details

The algorithms were implemented in *Python 3.4*, with a dependence on the *scipy* [1] library.

1.1 Maximum likelihood estimation

TODO

You need to choose proper parametric forms for the underlying probability distributions. In the report, you need to specify the parametric forms, the maximum-likelihood estimate for the parametric forms, and the resulting discriminant functions for classification for each dataset.

1.1.1 Normal Density

The discriminant function for the normal density is:

$$g_i(\mathbf{x}) = \mathbf{x}^T \mathbf{W}_i \mathbf{x} + \mathbf{w}_i^T \mathbf{x} + w_{i0}$$

where:

$$\mathbf{W}_i = -0.5 \mathbf{\Sigma}_i^{-1}$$

$$\mathbf{w}_i = \mathbf{\Sigma}_i^{-1} \mu_i$$

$$w_{i0} = -0.5 \mu_i^T \mathbf{\Sigma}_i^{-1} \mu_i - 0.5 \ln(\det \mathbf{\Sigma}_i) + \ln P(\omega_i)$$

Using the training samples, the mean and covariance can be estimated with maximum likelihood using the following definitions:

$$\hat{\mu}_i = \frac{1}{n} \sum_{k=1}^n \mathbf{x}_k \quad \hat{\mathbf{\Sigma}}_i = \frac{1}{n} \sum_{k=1}^n (\mathbf{x}_k - \hat{\mu}_i)(\mathbf{x}_k - \hat{\mu}_i)^T$$

Which yields:

$$p(\mathbf{x}|\omega_i) = \frac{1}{\sqrt{(2\pi)^n \det \hat{\mathbf{\Sigma}}_i}} \exp\left(-\frac{1}{2}(\mathbf{x} - \hat{\mu}_i)^T \hat{\mathbf{\Sigma}}_i^{-1}(\mathbf{x} - \hat{\mu}_i)\right)$$

The classification of instance \mathbf{x} is $\omega_i = \arg \max_{\omega_i} |p(\mathbf{x}|\omega_i)P(\omega_i)|$

1.1.2 Uniform

TODO

write out eqns

1.2 Parzen window estimation

TODO

You need to choose proper window functions, which you need to specify in the report along with the resulting discriminant functions for classification for each dataset. Here we choose the parameter values using the leave-one-out performance on the training set: For a set of candidate values, we compute the leave-one-out performance on the training set for each candidate and the optimal one is the one that gives the best leave-one-out performance (in case there are ties, specify how the ties will be broken).

1.3 Basic k-nearest neighbor

TODO

In this case, you need to implement the k-nearest neighbor classifier by first finding the first k nearest neighbors of an input and then classifying it as the class that appeared most among the nearest neighbors. Here you need to specify the distance measure between two inputs. Here we choose the parameter values using the leave-one-out performance on the training

set: For a set of candidate values, we compute the leave-one-out performance on the training set for each candidate and the optimal one is the one that gives the best leave-one-out performance (in case there are ties, specify how the ties will be broken).

2 Experimental Results

TODO

Document the classification results on the test set using each of the methods. For each dataset, you need to first estimate the necessary data normalization and parameters using the training set only (you need to document the results for different choices of parameters on the training set in the leave-one-out sense for Parzen windows and k -nearest neighbor rule)

2.1 Iris Data Set

2.1.1 Maximum likelihood estimation

1. **Normal Density** The estimated parameters of $\hat{\theta}$ for each of the classes from the training samples were:

$$\hat{\mu}_1 = (4.98181818, 3.39090909, 1.45151515, 0.25151515)$$

$$\hat{\Sigma}_1 = \begin{pmatrix} 0.10876033 & 0.08619835 & 0.02033058 & 0.01093664 \\ 0.08619835 & 0.13597796 & 0.01410468 & 0.00865014 \\ 0.02033058 & 0.01410468 & 0.03401286 & 0.00825528 \\ 0.01093664 & 0.00865014 & 0.00825528 & 0.0134068 \end{pmatrix}$$

$$\hat{\mu}_2 = (5.89090909, 2.78787879, 4.26363636, 1.31515152)$$

$$\hat{\Sigma}_2 = \begin{pmatrix} 0.2353719 & 0.06867769 & 0.165427 & 0.05256198 \\ 0.06867769 & 0.08530762 & 0.07743802 & 0.04442608 \\ 0.165427 & 0.07743802 & 0.20110193 & 0.06933884 \\ 0.05256198 & 0.04442608 & 0.06933884 & 0.0394674 \end{pmatrix}$$

$$\hat{\mu}_3 = (6.66060606, 2.94848485, 5.58484848, 1.99393939)$$

$$\hat{\Sigma}_3 = \begin{pmatrix} 0.36359963 & 0.11887971 & 0.2851607 & 0.03976125 \\ 0.11887971 & 0.10613407 & 0.08861341 & 0.03817264 \\ 0.2851607 & 0.08861341 & 0.29219467 & 0.03202938 \\ 0.03976125 & 0.03817264 & 0.03202938 & 0.06784206 \end{pmatrix}$$

This method correctly classified 48 of the 51 testing samples (94.1% accuracy).

TODO

add decision boundary graphs

2. Uniform

TODO

gotta do it

2.1.2 Parzen window estimation

TODO

document results for diff params on training set in the leaveone-out sense and document results.

2.1.3 Basic k-nearest neighbor

TODO

you need to first estimate the necessary data normalization and parameters using the training set only (you need to document the results for different choices of parameters on the training set in the leaveone-out sense for Parzen windows and k-nearest neighbor rule) and then document the classification results on the test set using each of the methods.

2.2 UCI Wine Data Set

TODO

you need to first estimate the necessary data normalization and parameters using the training set only (you need to document the results for different choices of parameters on the training set in the leaveone-out sense for Parzen windows and k-nearest neighbor rule) and then document the classification results on the test set using each of the methods.

2.3 Handwritten Digits Data Set

TODO

you need to first estimate the necessary data normalization and parameters using the training set only (you need to document the results for different choices of parameters on the training set in the leaveone-out sense for Parzen windows and k-nearest neighbor rule) and then document the classification results on the test set using each of the methods.

3 Analysis

TODO

You need to compare different methods in terms of classification performance and required time for classification, and give justifications for your observed empirical results.

4 Extra Credit

TODO

Please state clearly in your report if you have implemented any of the following extra credit options.

4.1 Speed up using k-d tree

TODO

With k from 1 to 10 with an increment of 1, first build a k -d tree from the training set and then classify the test samples using the k -nearest neighbor classifier by finding the nearest neighbors using the k -d tree. Compare the classification accuracy and the number of distance calculations with the basic k nearest neighbor implementation on the three datasets. Summarize your observations and justify your results.

4.2 Recognition of my handwritten digits

TODO

Apply the best classifier you have for hand written digit recognition on a test set consisting of your own written digits (you need to create the dataset). Document the classification performance, what you have done to improve the performance, and any additional issues you have handled.

References

- [1] Jones E, Oliphant E, Peterson P, *et al.* **SciPy: Open Source Scientific Tools for Python**, 2001-, <http://www.scipy.org/> [Online; accessed 2015-10-24].