Automatic Recognition of Pictured Dishes in Food-101

Abishek Kumar
Department of Computer Sciences
University of Wisconsin-Madison
akumar225@wisc.edu

Abstract

Classification in machine learning is a supervised learning approach in which a computer program learns from the data given to it and makes new predictions. This input data called the training data contains characteristics or features about the samples under observation along with a label for each of them. The core goal of classification is to learn a mapping between the input features and labels and use that mapping to predict the labels of new unseen data. In this project, we will address the problem of automatically recognizing pictured dishes on Food-101 dataset using three different approaches and evaluate their performance. The dataset consists of 101 food categories with 101,000 images, 1000 images for each category split into 750 training images, and 250 test images. The training data is left intentionally unclean and can contain noise in both features and labels. We will study the performance of Linear Regression, KNN, and Neural Networks on this dataset and propose a recommended approach based on our results.

1 Introduction

Classification is an important tool in today's world where big data is used to make all kinds of decisions in economics, medicine, and more. For example, detecting spam in emails can be identified as a classification problem with two classes as spam and not spam. The problem of object detection and classification is a widely researched topic in machine learning due to its broad range of applications. There is an array of different algorithms that have been developed over the years to solve this problem. Food classification in particular has many applications like helping patients track their calorie intake, auto-organizing of pictures in mobile phones, etc. But, unlike other image recognition tasks wherein we have definitive features separating each class, the problem of food classification is unique due to its high intra-class variability. Lighting, orientation, and the very realization of a recipe are some of the factors which can contribute to this variability.

In this project, we will try to understand the performance of simple classification algorithms like Linear Regression, KNN for the food classification task and compare their performance against more sophisticated algorithms like Neural Networks. We will use the Food-101 dataset for training our classifiers and evaluate their performance using holdout method, precision/recall, and ROC curve. Finally, we will propose a recommended algorithm based on our results.

2 Dataset: Food - 101

The dataset consists of 101 food categories with 101,000 images, 1000 images for each category split into 750 training images and 250 test images. Both the training and test images have been downloaded from foodspotting.com. The dataset has the top 101 food categories from the website. The test images have been manually cleaned, while the training images were left intentionally unclean and have some amount of noise. This noise comes mostly in the form of intense colors and sometimes mislabelled samples. All images have been re-scaled to have a maximum side length of 512 pixels and smaller images have been discarded.

3 Algorithms

3.1 Linear Regression

In statistics, linear regression is a method of modelling relationship between a scalar response (dependent variable) and one or more explanatory variables (independent variables). This relationship is modelled as a linear predictor function whose unknown parameters theta are estimated from the data. There are many flavors of Linear Regression like Least Squares Regression, Ridge Regression, Lasso Regression, etc. In this project, we will be using Linear Regression with Tikhonov regularization, also known as Ridge Regression for our classification. The motivation behind choosing Ridge Regression is due to its robustness to noise in the input dataset. Since our training data has noise in both features and labels, this method would give better performance compared to ordinary Least Squares approach.

3.2 K Nearest Neighbours

KNN is a type of lazy learning algorithm where labels are assigned to a new sample based on the labels of it's K closest neighbors. It's a simple algorithm where the training step involves just memorizing or storing the input features in an n-dimensional space. Classification involves plotting the new data point in the same n-dimensional space and choosing the class based on voting. The class which gets the most votes in the K Nearest Neighbors of the new sample will be assigned to the new sample.

3.3 Neural Networks

A neural network is a series of algorithms that endeavours to recognize the underlying relationships between data through a process that mimics the human brain. In recent times, neural networks have surpassed all previous benchmarks in the tasks of object recognition and classification. This is mainly due to the ability of neural networks to learn features of their own without any feature engineering. In this project, we will experiment with a neural network of 3 - 5 layers and understand how its performs compared to the other two algorithms

4 Milestones

- October 22nd: Initial project proposal.
- November 17th: Apply Ridge Regression for classification and produce results.
- November 29th: Apply KNN Algorithm for classification and produce results.
- December 7th: Develop a neural network for classification and produce results.
- December 14th: Compare the results between all three algorithms and prepare final report.

5 Final Deliverables

The implementation of each algorithm along with the final report will be available in this Github Repository - Automatic-Recognition-of-Pictured-Dishes-in-Food-101.

6 References

[1] Bossard, L., Guillaumin, M., & Van Gool, L. (2014) Food-101 – Mining Discriminative Components with Random Forests, *European Conference on Computer Vision*.

[2] S. Xiang, F. Nie, G. Meng, C. Pan & C. Zhang, "Discriminative Least Squares Regression for Multiclass Classification and Feature Selection," in IEEE Transactions on Neural Networks and Learning Systems, vol. 23, no. 11, pp. 1738-1754, Nov. 2012, doi: 10.1109/TNNLS.2012.2212721.