Project Part 1

Density Estimation and Classification

Prepared by **Rohit Roongta**

21st September 2019

NAÏVE BAYES CLASSIFICATION

Introduction

Naïve Bayes classification is a supervised learning algorithm used to classify discrete random output variable using Bayes theorem with an assumption of independent input variables.

Objective

To create a model to detect digit "7" and "8" using MNIST data subset which has 6265 and 5851 training images for "7" and "8" digits respectively and to test the accuracy using 1028 images of "7" digit and 974 images of "8" digit.

Below two features for each image need to be extracted:

- The average of all pixel values in the image
- The standard deviation of all pixel values in the image

Formula:

1) Mean (μ) = $\sum_{i=0}^{N} X$ / N; Here, N is number of pixels = 784 (28*28)
2) Variance (σ^2) = $\sum_{i=0}^{N}(X - \mu)^2$/ N Here, N is number of pixels = 784 (28*28)
3) Standard deviation (σ)= $\sqrt{Variance}$
4) Prior probability P(A) = Number of training set for A/Total number of training set
5) Covariance Matrix $\sum$ where x_{ij} = $(\sigma_i)(\sigma_j)$; i.e product of standard deviation of i^{th} feature and j^{th} feature

   Here, $\sum = \begin{matrix} (\sigma_1)^2 & 0 \\ 0 & (\sigma_2)^2 \end{matrix}$ as variables are assumed to be independent; diagonal elements will be equal to variance while remaining elements will be zero
6) Multivariate Normal Distribution $\rho(x)$ = $\frac{1}{(2\pi)^{\frac{d}{2}} * |\sum|^{\frac{1}{2}}}$ * exp [$-\frac{1}{2}$(x- μ)^t. $\sum^{-1}$. (x- μ)]; Here, d(no of features)= 2
   1) $|\sum|$ - determinant of covariance matrix
   2) $\sum^{-1}$ – Inverse of covariance matrix
7) Multivariate normal distribution for independent variables is equal to the product of univariate normal distribution of each variable.[1]

Pseudo Code (Algorithm)

- Feature extraction of training input set: average (feature 1) and standard deviation (feature 2) of all the pixels.
- Calculation of below variables for training input set:
  - Mean and standard deviation of features for all the labels.
  - Prior probability of all the labels.
  - Covariance, inverse of covariance and covariance determinant of the labels.

- For each input in testing set – features are extracted followed by calculation of the probability for digit 7 and digit 8 using the product of multivariate distribution and prior probability.

- The label with highest probability is the predicted output for the input set.

Variable Values

Digit 7

- Mean - [0.11452769775108732, 0.287556565177484]
- Standard deviation - [0.030632404696488404, 0.038201083694320306]
- Prior probability - 0.5170848464839881
- Covariance matrix -
  [[0.00093834 0.]
  [0. 0.00145932]]
- Inverse of covariance matrix -
  [[1065.70699895 0.]
  [0. 685.24935205]]
- Determinant of covariance matrix - 1.3693471065333262e-06

Digit 8

- Mean - [0.15015598189369694, 0.3204758364888717]
- Standard deviation - [0.038632488373958926, 0.03996007437065861]
- Prior probability - 0.4829151535160119
- Covariance matrix -
  [[0.00149247 0.]
  [0. 0.00159681]]
- Inverse of covariance matrix -
  [[670.03059639 0.]
  [0. 626.24954644]]
- Determinant of covariance matrix -2.3831860101893997e-06

Accuracy Result:

DIGIT 7:

Result: 781 out of 1028
Accuracy: 75.9727626459144

DIGIT 8:
Result: 611 out of 974
Accuracy: 62.73100616016427

TOTAL RESULT: 1392 out of 2002
TOTAL ACCURACY: 69.53046953046953

RUNTIME DURATION: 9.609917 seconds

Reference Link:

[1] - http://cs229.stanford.edu/section/gaussians.pdf

LOGISTIC REGRESSION

Introduction

Logistic regression is statistical model that predicts discrete values using sigmoid logistic function and a cut off threshold value.

Objective

To create a model to detect digit "7" and "8" using MNIST data subset which has 6265 and 5851 training images for "7" and "8" digits respectively and to test the accuracy using 1028 images of "7" digit and 974 images of "8" digit.
Below two features for each image need to be extracted:

- The average of all pixel values in the image
- The standard deviation of all pixel values in the image

Formula:

1) Mean (μ) = $\sum_{i=0}^{n} X$ / N
2) Variance (σ^2) = $\sum_{i=0}^{n}(X - \mu)^2$/ N
3) Standard deviation (σ)= $\sqrt{Variance}$
4) Sigmoid logistic function = $\frac{1}{1+\exp(-\theta^t X)}$ where θ is the parameter vector and X is input vector
   Here, $\theta^t X = \theta_0 x_0 + \theta_1 x_1 + \theta_2 x_2$ where $x_0 = 1$ (also known as bias)
5) Gradient ascent:
   Until θ converage:
   Error function = (Expected value – Predicted Value) = (Y – sigmoidal function)
   Gradient = Product of training input vector and error function
   Parameter = parameter + α*gradient [α is the learning rate]

Pseudo Code (Algorithm)

- Feature extraction of training input set: average (feature 1) and standard deviation (feature 2) of all the pixels, and insertion of 1s column at the start in the input vector which would represent $\boldsymbol{x_0}$-bias – augmented vector
- Set $\boldsymbol{\theta}$ to some arbitrary and calculate the final value of $\boldsymbol{\theta}$ using gradient ascent with $\boldsymbol{\alpha}$=0.001
- For each input in testing set – features are extracted followed by calculation of the probability using new parameter in sigmoidal function.
    - If the resultant probability is greater than 0.5 (threshold) – it is digit 8 (label 1)
    - Otherwise it is digit 7 (label 0)

Variable Values

DEFAULT VALUE OF PARAMETERS $(\theta_0, \theta_1, \theta_2)$
[[1.]
[1.]
[1.]]

FINAL VALUE OF PARAMETERS $(\theta_0, \theta_1, \theta_2)$
[[22.72233365]
[243.85233715]
[-177.90606678]]

Total number of iterations: 80000 iterations with learning rate = 0.001

Accuracy Result:

DIGIT 7:
Result: 787 out of 1028
Accuracy: 76.55642023346303

DIGIT 8:
Result: 855 out of 974
Accuracy: 87.782340862423

TOTAL RESULT: 1642 out of 2002
TOTAL ACCURACY: 82.01798201798202

RUNTIME DURATION: 32.959579 seconds