CSE -575 SPRING 2020

Project Part 1: Density Estimation and Classification

Name: Guru Preetam Kadiri

ASU ID: 1217291777

Azurite ID: gkadiri

1. Introduction

Aim of the project is to perform parameter estimation for the given dataset using generative – Naïve Bayes, and discriminative – Logistic Regression methods. We use the given MNIST dataset which contains 70,000 images of handwritten digits. The dataset is divided into 60,000 training images and 10,000 testing images. The images for digit ‘7’ and ‘8’ have been used for the scope of this project.

Statistics of the given dataset

Class	Training	Testing
7	6265	1028
8	5851	974
Total	12116	2002

2. Feature Extraction

The images contain grey levels as a result of anti-aliasing techniques. They are a 28x28 images, i.e. a total of 784 pixels and it is stored in a 784-dimensional array. We are extracting the mean and standard deviation from these pixel values to not use the whole array as features, but instead use 2-dimensional arrays. We assume that the 2 features are independent.

Notations Used:

training_X – Training samples

training_Y – Training labels

testing_X – Testing Samples

testing_Y - Testing labels

mean_training_X – mean

sd_training_X – standard deviation

no_samples – Number of training samples

classes – Number of classes: 2 – 7 and 6

features – Number of features

Maximum Likelihood Estimation

$$P(x|\ \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\ \mu)^2}{2\sigma^2}\right)$$

Maximum Likelihood estimation can be used to estimate the parameters (μ, σ), which will be used in Naïve Bayes. We construct a multivariate normal distribution as the feature vectors are 2D. The normal distribution estimates the Probability Density Function (PDF) for the given samples.

Apply log-likelihood and zeroing its gradient gives the mean

$$\log P(x \mid \mu, \sigma) = -\left(\frac{1}{2}\log|\sigma| + \frac{1}{2}(x-y)^T \sigma^{-1}(x-\ \mu)\right)$$

Which gives

$$\mu = \frac{1}{n}\sum_{i=1}^{n}(x^i)$$

$$\sigma = \frac{1}{N-1}\sum_{i=1}^{n}(x^i - \mu)(x^i -\ \mu)^T$$

The covariance matrix σ, is calculated using μ. Covariance matrix is a 2x2 matrix.

The Maximum Likelihood Function is defined in MLE()

For class 7:

μ

Mean for class 7:-

[[0.1145277]

[0.28755657]]

σ

covaraiance matrix for class 7:-

[[0.00048524 0.0005967]

[0.0005967 0.00075466]]

For class 8:

μ

Mean for class 8:-

[[0.15015598]

[0.32047584]]

σ

Covariance matrix for class 8:-

[[0.0007208 0.0007309]

[0.0007309 0.00077119]]

3. Naïve Bayes Classification

Using Naïve Bayes, we can calculate the PDF. The algorithm predicts if the input belongs in the class or not.

$$P(Y|X) = \frac{P(Y)P(X|Y)}{P(X)}$$

$$P(Y|X) \propto P(y)\prod P\left(x^i|\mu,\sigma\right)$$

P(y) – prior ability;

P($y|\mu,\sigma$) can be calculates using P(y) and P(x$|\mu,\sigma$)[given by guassian distribution].

Log Likelihood to generate Naives Bayesian.

$$\log\left(P(y|\mu,\sigma) = -(\log\left(\sigma^{\frac{1}{2}}\right) + \frac{1}{2}(x - \mu^T)\sigma^{-1}(x-\mu)\right.$$

$$Prediction = argmax(\log\left(P(y|\mu,\sigma)\right.$$

For Class 7:

probability(y=7): 0.5170848464839881

Diagonal Covariance Matrix:

[[0.00048524 0.   ]

[0.   0.00075466]]

For Class 8:

probability(y=8): 0.4829151535160119

Diagonal Covariance Matrix:

[[0.0007208 0.   ]

[0.   0.00077119]]

4. Logistic Regression

Logistic regression is a linear classifier, sigmoid function is used to predict. P(Y|X) is modelled using logit function.

The logit function

$$P(Y|X) = \frac{1}{1 + e^{(w_0 + w_1 x + w_2 x)}}$$

To achieve we use the following formula Likelihood function

$$\sigma(w^t x) = \frac{1}{1 + e^{-(w^T x)}}$$

$$P(Y|X) = (\sigma(w^T))^y + (1 - y)\ (1 - \sigma(w^T x))^{1-y}$$

Gradient is found to maximize likelihood function

$$L(w) = \log\big(P(Y|X)\big) = ylog\big(\sigma(W^T x)\big) + (1 - y)\log\ (1 - \sigma(W^T x))$$

$$\nabla L(w) = \sum [y - \sigma(W^T x)] x^i$$

Gradient = error * x

Error = actual_y – predicted_y

Gradient ascent is applied using Learning_Rate(β)

$$w^{(kn)} = w^k + \beta \nabla_w L(w^k)$$

K - classes

The gradients are calculated by keep

ITERATIONS: 10000

LEARNING RATE: .00001

Weights are predicted, when the gradient ascent converges

$$if\ \sigma^Y w^T \leq 0\ impliies\ y = 0;\ \sigma^Y w^T \geq 0\ impliies\ y = 0$$

Finally

The trained weights are

[[-0.07167472]

[0.03933775]

[-0.08432989]]

5. Results

Accuracy	Total Samples	Samples for class 7	Samples for class 8
Naïve Bayes Classifier	68.73126873126873	71.40077821011673	65.91375770020534
Logistic Regression	49.8001998001998	69.26070038910505	29.260780287474333