Homework

Machine Learning

Sandeep Saurav

203310021

GNR-652

Data sets :-

BankNote Authentication :-

* Features (No. of Attributes) -
  1. Variance of Wavelet Transformed Image (VoWTI)
  2. Skewness of Wavelet Transformed Image (SoWTI)
  3. Curtosis of Wavelet Transformed Image (CoWTI)
  4. Entropy of Image (EOI)
  5. Class.
     - 1 → Authentic Note
     - 0 → Duplicate Note.

Training Set :-

i_{train}	VoWTI (x_1)	SoWTI (x_2)	CoWTI (x_3)	EOI (x_4)	Class (Y)
1.	3.6216	8.6661	−2.8073	−0.44699	0
2.	4.5459	8.1674	−2.4586	−1.4621	0
3.	3.866	−2.6383	1.9242	0.10645	0
4.	3.4566	9.5228	−4.0112	−3.5944	0
5.	0.32924	−4.4552	4.5718	−0.9888	0
6.	4.3684	9.6718	−3.9606	−3.1625	0
7.	3.5912	3.0129	0.72888	0.56421	0
8.	2.0922	−6.81	8.4636	−0.60216	0
9.	3.2032	5.7588	−0.75345	−0.61251	0
10.	1.5356	9.1772	−2.2718	−0.73535	0
11.	−1.3971	3.3191	−1.3927	−1.9948	0
12.	0.39012	−0.14279	−0.031994	0.35084	0
13.	−1.6677	−7.1535	7.8929	0.91765	1
14.	−3.8483	−12.8047	15.6824	−1.281	1
15.	−3.5681	−8.213	10.083	0.96765	1
16.	−2.2804	−0.30626	1.3347	1.3763	1
17.	−1.7582	2.7397	−2.5323	−2.234	1
18.	−0.89409	3.1991	−1.8219	−2.9452	1
19.	0.3434	0.12415	−0.28733	0.14654	1
20.	−0.9854	−6.661	5.8245	0.5461	1
⋮	⋮	⋮	⋮	⋮	⋮

Test set :-

i_{test}	VoWTI (x_1')	SoWTI (x_2')	CoWTI (x_3')	EOI (x_4')	Class (y')
1.	2.6606	3.1681	1.9619	0.18662	0
2.	−1.3887	−4.8773	6.4774	0.34179	1
3.	3.931	1.8541	−0.023425	1.2314	0
4.	−3.7503	−13.4586	17.5932	2.7771	1
5.	0.01727	8.693	1.3989	−3.9668	0
6.	0.40614	1.3492	−1.4501	−0.55949	1
7.	3.2414	0.40971	1.4015	1.1952	0
8.	−3.5637	−8.3827	12.393	−1.2823	1
9.	2.2504	3.5757	0.35273	0.2836	0
10.	−2.5419	−0.65804	2.6842	1.1952	1

Normalized Test set using :- $Z = \frac{X_{test} - \mu}{\sigma}$

$\mu = [0.4522 \quad 0.5538 \quad 2.1753 \quad -0.1972]$

$\sigma = [2.763 \quad 6.8225 \quad 4.9574 \quad 1.784]$

X_{test} :-

i_{test}	VoWIT (x_1')	SoWIT (x_2')	CoWIT (x_3')	EOI (x_4')	
~~1~~ 1.	0.[illegible]	0.382	−0.043	0.6411	$\rightarrow$ 0
~~7~~ 7.	1.2590	0.190	−0.4435	1.22	$\rightarrow$ 0
~~2~~ 2.	−1.1574	1.192	−0.1566	−1.667	$\rightarrow$ 0
~~7~~ 8.	1.0	−0.0211	−0.1560	1.206	$\rightarrow$ 0
~~3~~ 3.	0.650	0.4429	−0.3676	0.695	$\rightarrow$ 0
~~6~~ 6.	−0.0167	0.1165	−0.7313	0.223	$\rightarrow$ 1
~~7~~ 4.	−0.666	−0.7960	0.867	0.728	$\rightarrow$ 1
~~8~~ 9.	−1.520	−2.0538	3.11	−1.02	$\rightarrow$ 1
~~9~~ 5.	−1.453	−1.3098	2.06	−0.1822	$\rightarrow$ 1
~~9~~ 10.	−1.0836	−0.1776	0.10	1.206	$\rightarrow$ 1

1) Using KNN (k=5)

$$d_{ij} = \sqrt{(x_1 - x_1')^2 + (x_2 - x_2')^2 + (x_3 - x_3')^2 + (x_4 - x_4')^2}$$

j_{test}	K-Nearest (d_{ij})	Class Y_{train}	Class Y_{pred} test
1.	(0.39, 0.469, 0.826, 0.85, 0.95)	(0, 0, 0, 0, 0)	0
7.	(0.45, 0.61, 0.85, 0.99, 1.08)	(0, 0, 0, 0, 0)	0
2.	(0.88, 0.90, 1.22, 1.42, 1.55	(0, 0, 1, 1, 1)	1
8.	(0.59, 0.88, 0.597, 0.55, 0.79)	(0, 0, 0, 0, 0)	0
3.	(0.24, 0.52, 0.72, 0.86, 0.87)	(0, 0, 0, 1, 1)	0
6.	(0, 0.76, 0.76, 0.76, 0.76)	(1, 1, 1, 1, 1)	1
4.	(0, 0.33, 0.42, 0.34, 0.85)	(1, 0, 0, 0, 0)	0
9.	(0, 0.9, 1.56, 0.92, 2.19)	(1, 0, 1, 0, 1)	1
5.	(0, 0.82, 0.93, 0.99, 1.7)	(1, 0, 0, 1, 1)	1
10.	(0, 0.66, 1.0, 1.35, 1.35)	(1, 0, 0, 1, 1)	1.

$$\text{Accuracy} = \frac{8}{10} = 0.8.$$

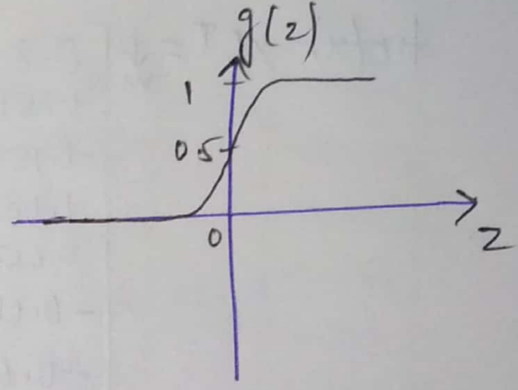
⇒ Accuracy using K=4 NN, using 10 sample test comes out to be 80%.

2) Using Logistic Regression

$$h_\theta(x) = g(\theta_0 + x_1\theta_1 + x_2\theta_2 + \dots + x_n\theta_n) = g(X\theta^T)$$

$x_0 = 1$ bias

$$\underset{m\times 1}{[h_\theta(x)]} = \left[g\left(\underset{m\times(n+1)}{[X]}\underset{(n+1)\times 1}{[\theta^T]}\right)\right]_{m\times 1}$$



where, $g(z) = \dfrac{1}{1+e^{-z}}$ } sigmoid fn.

$0 \le g(z) \le 1$

$y_{pred} = 1$ if $h_\theta(x) \ge 0.5$ i.e, when $X\theta^T \ge 0$
(or) $g(X\theta^T) \ge 0.5$)

$y_{pred} = 0$ if $h_\theta(x) < 0.5$ i.e, when $X\theta^T < 0$
(or) $g(X\theta^T) < 0.5$)

Cost fn $J(\theta) = \frac{1}{m}\left(y^T \log(h) - (1-y)^T \log(1-h)\right)$

Cost Min. algo. $\longrightarrow$ Gradient Descent (Stochastic G.D.)
$\longrightarrow$ θ_{old} = [Some initial random small no.]

Repeat {
  for $i = 1$ to m
  { $\theta_{j\,new} := \theta_{j\,new} - \frac{\alpha}{m}\left(h_\theta(x^{(i)}) - y^{(i)}\right)x_j^{(i)}$
  }
} for $j = 1$ to n.

$m \to$ training sample
$n \to$ feature vectors.
$\alpha = 0.1 \to$ learning Rate Parameter.

Step 1:- finding weight vector using Training data set.

Let $\theta_{old} = [0.672 \quad 0.488 \quad 0.825 \quad 0.0314 \quad 0.808]$

$\theta_{new} = [0.658 \quad 0.4284 \quad 0.786 \quad 0.061 \quad 0.801]$

, after many iteration diff. is very small & it gets converged.

$$\theta_{new} = \theta_{final} = \begin{array}{ccccc} [-2.785 & -13.013 & -40.793 & -7.896 & 3.308] \\ \theta_0 & \theta_1 & \theta_2 & \theta_3 & \theta_4 \end{array}$$

$$h_\theta(x) = g(\underbrace{-2.785 - 13.0134x_1' - 4.793x_2' - 7.8963x_3' + 3.3084x_4'}_{X\theta^T})$$

$\downarrow$ Bias $x_0 = 1$

$$\underset{\uparrow \; X_{test}}{X\theta^T} = \begin{bmatrix} 1 & 0.8 & 0.383 & -0.043 & 0.6411 \\ 1 & 1.259 & 0.190 & -0.04435 & 0.22 \\ 1 & -0.1574 & 1.192 & -0.1566 & -1.687 \\ 1 & 1 & -0.0211 & -0.1560 & 1.206 \\ 1 & 0.65 & 0.04129 & -0.3676 & 0.695 \\ 1 & -0.0167 & 0.1165 & -0.7313 & 0.223 \\ 1 & -0.666 & -0.7960 & 0.8677 & 0.728 \\ 1 & -1.52 & -2.0538 & 3.110 & -1.02 \\ 1 & -1.453 & -1.3098 & 2.061 & -0.182 \\ 1 & -1.0836 & -0.1776 & 0.1026 & 1.206 \end{bmatrix} \begin{bmatrix} -2.785 \\ -13.0134 \\ -4.793 \\ -7.8963 \\ 3.3084 \end{bmatrix} = \begin{bmatrix} -12.561 \\ -12.5219 \\ -10.7989 \\ -10.596 \\ -8.173 \\ 3.38 \\ 5.257 \\ -1.080 \\ 5.53 \\ 15.349 \end{bmatrix}$$

$X\theta^T$	$h_\theta(x)$	Y_{actual}	$Y_{predict}^{test}$
-12.561	0	0	0
-12.5219	0	0	0
-10.7989	0	0	0
-10.596	0	0	0
-8.173	0	0	0
3.38	1	1	1
5.257	1	1	1
-1.080	0	1	0
5.53	1	1	1
15.349	1	1	1

% Accuracy Based on 10 test Sample $= \frac{9}{10} = 0.9 \simeq 90\%$.