TANMAY GUPTA
tg 289

MATH-680

HW-2

(1)

Paper -> Choosing btw Logistic Regression & Discriminant Analysis

-> From the paper we can understand that logistic regression with max. likelihood outperforms Discriminant Analysis in the problem of nonnormality.

-> They have shown it with the help of 2 examples. first of breast cancer patients which is a classification problem with 0 as the lymph nodes not involved and 1 as involved. LR classified 82 of the 115 patients whereas LDA classified only 77 with classification rate of 71.30 and 62.07 respectively. Thus LR outperforms LDA in the example.

-> The second example is of population change in 50 states in the US in which LR gives a correct classification of 80% as compared to LDA with 68%. Thus we can see it performs better.

(2)

(a)

Radius of a uniformly dis. ball i.e sphere is 1.

$$P(x \in \text{sphere}(r)) = \frac{S(p)\, r^p}{S(p)\, 1^p}$$

$$= r^p$$

The probability of a point falling in a sphere is proportional to its volume.

Euclidean distance $\|x\|$ of ~~any~~ a p-dim. vector x

$$P(x \in S_p(r)) = P(\|x\| \leq r)$$

$$= r^p$$

$$R = \min \|x_i\|.$$

$$P(R \leq r) = P_r(\min \|x_i\| \leq r)$$

$$= 1 - P_r(\min \|x_i\| > r)$$

$$= 1 - \prod_{i=1}^{n} P(\|x_i\| > r)$$

$$= 1 - \prod_{i=1}^{n} (1 - P(\|x_i\| \leq r))$$

$$= 1 - (1 - r^p)^n$$

Median is at 1/2.

$$1-(1-r_{1/2}^{p})^{b} = 1/2$$

$$r = \left(1 - \frac{1}{2}^{1/b}\right)^{1/p}$$

(b) If $b = 500$ & $p = 10$.

$$r_{1} = \left(1 - \frac{1}{2}^{1/500}\right)^{1/10}$$

$$= \left(1 - 0.5^{0.002}\right)^{0.1}$$

$$= \boxed{0.52}$$

If $b = 500$ & $p = 100$

$$r_{1} = \left(1 - \frac{1}{2}^{1/500}\right)^{1/100}$$

$$= \left(1 - \frac{1}{2}^{0.002}\right)^{0.01}$$

$$= \boxed{0.94}$$

If p is large we get a larger value of r.

3

(a)

For a linear regression model ->

Training error -> 0.225

Testing error -> 0.22

(b)

For LDA

Training error : 0.225

Testing error -> 0.220

(C) For log regression

Training error : 0.29

Testing error = 0.23

For Bayes →

Training error = 0.285

Testing error = 0.219

The training & testing errors for linear reg & LDA are same and they are lower than logistic reg. & Bayes dec. boundary.

(4)

In R code

(5)

Training error = ~~0.255~~ 0.31

Testing error = ~~0.31~~ 0.285

Scanned by CamScanner