

علیرضا درویش 96109674 - تمرین 5 - تحلیل رگرسیون

4.

الف ، $\eta \approx 0.1$

ب ، $\eta \approx 0.1^2 = 0.01$

پ ، $\eta \approx 0.1^{100} = 10^{-100}$

ت ، در P های بزرگ ، درصد بسیار کمی از داده ها مورد استفاده قرار می گیرند

لیس KNN تقییبی خوبی نخواهد داشت

ث ، $a_{P=2} = 0.1$

$$a_{P=2} = \sqrt{0.1}$$

$$a_{P=100} = 0.1^{\frac{1}{100}} \approx 0.977$$

برای آینه در $P=100$ از 10٪ از داده ها استفاده کنیم ، باید از داده های بسیار دور

از نقطه ی $test$ استفاده کنیم . که این از دقت یلینی بیی می گذرد.

5.

الف) QDA روی train نتیجه بهتری دارد چون درجه آزادی

بیشتری دارد. اما LDA روی test نتیجه بهتری دارد چون QDA

روی train دچار overfit می شود.

ب) QDA روی train نتیجه بهتری دارد چون درجه آزادی

بیشتری دارد. در مورد test، به مرز Bayes بستگی دارد. اگر مرز

Bayes درم 2 باشد، QDA بسیار نتیجه خوبی دارد. اما اگر

مرز بسیار غیرخطی باشد، نتیجه هم LDA و QDA شبیه به هم است و بی

QDA احتمالاً اندکی بهتر است.

پ) در حالت کلی و با افزایش n ، پارامترهای QDA دقت بیشتری به

دست می آیند. چون تعداد پارامترهای QDA خیلی بیشتر از LDA است

پس دقت QDA افزایش می یابد.

لست به LDA

JAHAN NAMA

ت، ~~هو~~ غلط. خطای QDA بی test للشراف

چون اگر مرد واقعی خطی باشد، QDA بی train ، overfit می کند.

$$E(X|Yes) = 10, E(X|No) = 0 \quad .7$$

$$\hat{\sigma}^2 = 36 \Rightarrow \hat{\sigma} = 6$$

$$P(Yes|X=4) = \frac{P(X=4|Yes) P(Yes)}{P(X=4)}$$

$$= \frac{P(X=4|Yes) P(Yes)}{P(X=4|Yes) P(Yes) + P(X=4|No) P(No)}$$

$$= \frac{e^{-\frac{(4-10)^2}{2 \times 36}} \times 0.8}{e^{-\frac{(4-10)^2}{2 \times 36}} \times 0.8 + e^{-\frac{(4-0)^2}{2 \times 36}} \times 0.2} = 0.75$$

8. انتی فیت من logistic reg است چون error-rate

2

3 برای KNN با $K=1$ برای داده های train، صفر است.

4

5 پس برای داده های test، 100 است که خیلی بیشتر از 30%

6

7 میل ادل است.

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24