**Answers - Classification - Assignment 3**

1. Compare and contrast Logistic regression with any probabilistic generative model using an example data set. **[6M**]

Answer1.

Logistic Regression is a Probabilistic discriminative approach, where the model discriminates input based on maximum likelihood of input and class in the training data.

Logistic regression models the Probability of target using a non-linear transformation applied on a linear function of the form z = w0 + w1\*x1 + W2\*x2 …., where W0 is the bias and w1..N are coefficients.

Logistic Regression can be applied on binary or multi-class classification problem, where each of the target class probabilities is represented using a linear function of input features.

Logistic Regression is best suited when the input data and target class is non linearly separable.

Performance of Logistic regression is not dependent on conditional independence of input features, and demonstrates Highest accuracy.

Whereas,

Probabilistic generative techniques models class conditional probability e.g. P(y|x) using joint probability distribution p(x,y), where P(x,y) = P(x|y) \* P(y)

Generative model uses P(x,Y) and applies Bayes Theorem to model the conditional probability such as P(Y|x) = P(x, Y) / P(Y).

Greatly depends on conditional independence of features and the target class distribution.

Like Logistic regression, generative model can be applied to multi class classification problems.

Let’s take an example of classifying if its sunny, will we play tennis

Sr. No. Temp Play Tennis

1. 27 yes

2. 28 Yes

3. 29 Yes

4. 34 Yes

5. 38 No

6. 31 Yes

7. 30 No

8. 39 No

9. 40 No

10. 16 yes

11. 18 No

12. 5 No

13. 4 No

14. 30 No

15. 30 Yes

16. 30 Yes

17. 30 Yes

**Using generative model**,

*Class = Yes / No,*

*P(Play=?|Temp = 30) = ?*

*P(Play=yes) = 6/17*

*P(Play=no) = 8/17*

*P(Play=Yes| Temp=30)*

*= For all X P(x|Y) / P(Y)*

*= (3/9 ) / 6/17*

*= .942*

*P(Play=No | Temp=30)*

*= ( 2/8 ) / 8/17*

*= .5313*

*As per a generative model the probability of Playing is .942, hence,*

*P(Play=?|Temp = 30) =* ***Yes***

**Using discriminative model**,

P(Play=yes | temp=30) = exp(y) / (exp(y) + 1, where

Y = w0 + w1 \* Xtemp

For simplicity, let’s assume, W0 = -8.3 , W1 = 0.612. In actual case, these coefficient are derived by minimising Cross Entropy error between the actual probability and predicted probability values.

P(play=yes| temp=30) = .999 if you evaluate,

= exp (-8.3+0.612 \* 30) / exp (-8.3+0.612 \* 30) + 1

= .999 => **YES**

P(play=yes| temp=4) = .002 if you evaluate,

= exp (-8.3+0.612 \* 4) / exp (-8.3+0.612 \* 4) + 1

= .002 => **NO**

1. For the dataset -2 (liver disease dataset), implement Logistic Regression classifier using Python. **[12M**]

Answer2.

See the attached python notebook

——————————————————————————————————

Submission Details

Text answers – id\_classification\_assignment3.doc

Code - id\_LR.ipynb