MT22139

3.a) Logistic Regression.

For N-class instances, how many binary classifier models are required to be generated for

- One vs All multi-class classification using logistic Regression
- One vs One multi-class classification using logistic regression

A. → One vs All is one of the method of multi-class classification. The kth classifier will classify the data as whether the data belongs to the kth class or not. We require N classifiers for N classes where each classifier predicts if the sample (input) belongs to that particular class or not

→ One vs one multi-class classification approach, we require $^N C_2$ classifiers i.e., $N(N-1)/2$ classifiers.
It splits the dataset into one dataset for each class versus every other class. as (N - No. of classes).

4.a) Prove that Gamma distribution belongs to the same family as curves as poisson Distribution. Be informed that both Gaussian and Bernoulli distributions also belong to this family.

A. We know,

poisson distribution is given by,

$$f_X(x) = \frac{e^{-\lambda} \cdot \lambda^x}{x!}$$

It can be written as $e^{\ln(\lambda^x)} \cdot e^{-\lambda} \cdot e^{-\ln(x!)}$

$$= e^{x \ln(\lambda)} \cdot e^{-\lambda} \cdot e^{-\ln(x!)}$$

Compare above equation with

$$\exp\left[\frac{x\theta - b(\theta)}{a(\phi)} + c(x,\phi)\right].$$

Standard probability distribution of exponential family.

$$x\theta = x\ln\lambda \Rightarrow \theta = \ln\lambda$$

$$b(\theta) = \lambda \ ; \ a(\phi) = 1$$

$$c(x,\phi) = -\ln(x!)$$

Hence, poission belongs to exponential family.

Proving Gamma distribution belongs to same family of curves as of poisson's distribution i.e, exponential family

Gamma distribution is given by,

$$f_X(x) = \begin{cases} \dfrac{\lambda^{\alpha}\, x^{\alpha-1}\, e^{-\lambda x}}{\Gamma(\alpha)} & x > 0 \\ 0 & \text{otherwise.} \end{cases}$$

For $x>0$, $f_X(x)$ can be written as.

$$f_X(x) = e^{-\lambda x}\, e^{\ln\left(\frac{\lambda^{\alpha} x^{\alpha-1}}{\Gamma(\alpha)}\right)}$$

$$= e^{-\lambda x} \cdot e^{\ln(\lambda^{\alpha})} \cdot e^{\ln(x^{\alpha-1})} \cdot e^{-\Gamma(\alpha)}$$

$$= e^{-\lambda x + \alpha\ln(\lambda) + (\alpha-1)\ln(x) - \Gamma(\alpha)}$$

$$= \exp\left[\frac{\frac{\lambda}{\alpha}x - \ln(\lambda)}{-1/\alpha} + (\alpha-1)\ln(x) - \Gamma(\alpha)\right].$$

Comparing the equation with standard equation

$$f_X(x, \theta, \phi) = \exp\left[\frac{x\theta - b(\theta)}{a(\phi)} + c(x, \phi)\right].$$

We obtain

$$\theta = \frac{\lambda}{\alpha} ; \quad a(\phi) = \frac{-1}{\alpha} ; \quad b(\theta) = \ln(\lambda);$$

$$c(x, \phi) = \left[(\alpha - 1)\ln(x) - \Gamma(\alpha)\right].$$

Hence, gamma distribution belongs to the exponential family of curves.

7.a) Derive the formula of F5 score in terms of precision and Recall in concurrence with F1 score. What will be the corresponding value of alpha in that case?

Ans

F1 score can be represented as Harmonic mean of precision and recall

$$\frac{2}{F1} = \left[\frac{1}{\text{precision}} + \frac{1}{\text{recall}}\right]$$

Considering the concept of weighted Harmonic mean:

$$\frac{1}{F\beta} = \alpha\left(\frac{1}{\text{precision}}\right) + (1-\alpha)\left(\frac{1}{\text{recall}}\right)$$

α = weight of precision.

$(1-\alpha)$ = weight of recall.

Let, precision is represented by P.
Recall is represented by R

$$\frac{1}{F_\beta} = \frac{\alpha R + (1-\alpha) P}{PR}$$

$$\Rightarrow F\text{-}\beta = \frac{PR}{\alpha R + (1-\alpha) P}$$

In order to find α, we find relative importance of P/R ratio using

$$\frac{\partial F\beta}{\partial R} = \frac{\partial F\beta}{\partial P}$$

$$\frac{\partial (F\beta)}{\partial R} = \frac{(\alpha R + (1-\alpha)P)P - PR(\alpha)}{(\alpha R + (1-\alpha)P)^2} \quad \text{—— ①}$$

$$\frac{\partial F\beta}{\partial P} = \frac{(\alpha R + (1-\alpha)P)R - PR(1-\alpha)}{(\alpha R + (1-\alpha)P)^2} \quad \text{—— ②}$$

Equating ① & ②.

$$\frac{\alpha R + (1-\alpha)P - PR(\alpha)}{(\alpha R + (1-\alpha)P)^2} = \frac{(\alpha R + (1-\alpha)P)R - PR(1-\alpha)}{(\alpha R + (1-\alpha)P)^2}$$

$$\alpha RP + P^2 - \alpha P^2 - PR\alpha = \alpha RR + PR - \alpha PR - PR + \alpha PR.$$

$$P^2 - \alpha P^2 = \alpha R^2$$

$$(1-\alpha) P^2 = \alpha R^2 \quad \text{—— ③}$$

By definition; $\beta = \frac{R}{P} \Rightarrow R = \beta P$.

Equation ③ becomes.

$$(1-\alpha)P^2 = \alpha(\beta P)^2$$

$$(1-\alpha)P^2 = \alpha\beta^2 P^2$$

$$P^2 - \alpha P^2 = \alpha\beta^2 P^2$$

$$P^2 = \alpha(\beta^2 P^2 + P^2)$$

$$P^2 = \alpha P^2(\beta^2+1)$$

$$\alpha = \frac{1}{\beta^2+1}$$

We had derived

$$F\beta = \frac{PR}{\alpha R + (1-\alpha)P}$$

$$= \frac{PR}{\left(\frac{1}{1+\beta^2}\right)R + \left(1 - \frac{1}{1+\beta^2}\right)P}$$

$$\therefore F\beta = \frac{PR(\beta^2+1)}{R+\beta^2 P}$$

F5 score i.e., $\beta = 5$

$$F5 = \frac{PR(25+1)}{25P+R} = \frac{26PR}{R+25P}$$

$$\boxed{F5 = \frac{26PR}{R+25P}}$$

(b) We took $\beta = 5$ i.e., $\beta > 1$.
Hence emphasis on recall is more than than that on precision.

5.a) Disadvantages / Drawbacks of K-Mean clustering Algorithm.

→ We need to specify the value of K (the number of clusters) in the beginning.

→ K-mean can only handle numerical data and the results might be skewed if we do not normalize it.

→ It cannot ~~cluster~~ be used for clusters with non-convex shapes.