_/_/_

Name → Ritisha Gupta  ML-ASS 2

Roll No. → MT22056

Ques 3) a) N-binary classifier models are required for / to be generated for one v/s all multi-class classification using logistic regression.

b) N*(N-1)/2 binary classifier models are required to be generated for one v/s one multi-class classification using logistic regression.

Ques 7) a) Derive F5 score

As we know that F1 score is the harmonic mean of precision and recall.

$$F_1 = \frac{2}{\frac{1}{recall} + \frac{1}{prec^n}} = \frac{2(precision)(recall)}{precision + recall} = \frac{2PR}{P+R}$$

where precision (P) $= \frac{TP}{TP+FP}$ & recall (R) $= \frac{TP}{TP+FN}$.

# In more general form, $F_\beta = \frac{(\beta^2+1)PR}{\beta^2 . P + R}$ where $(0 \leq \beta \leq \infty)$

& according to van Rijsbergen's E (effectiveness function), the def^n of F-measure

$$= \boxed{E = 1 - \frac{1}{\frac{\alpha}{P} + \frac{(1-\alpha)}{R}}, \text{ where } \alpha = \frac{1}{1+\beta^2}}$$

Removing $\alpha$ using $\beta$.

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

_/_/_

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

$$\boxed{E = 1 - \frac{(1+\beta^2)PR}{R+P\beta^2}}$$ → This part matches with our general $F_\beta$ formula.

so $\boxed{E = 1 - F_\beta}$

Hence we can also write $\boxed{F_\beta = \frac{1}{\frac{\alpha}{P} + \frac{(1-\alpha)}{R}}}$ $(0 \le \alpha \le 1)$ — (1)

~~For par~~

So using $\boxed{F_\beta = \frac{(1+\beta^2)PR}{R+P\beta^2}}$ put $\beta = 5$ to get $F_5$ score

$$F_5 = \frac{(1+(5)^2)PR}{R+25P} = \frac{26PR}{R+25P}$$

So $\boxed{F_5 = \frac{26PR}{R+25P}}$

# Finding value of $\alpha$ ⇒ $\alpha = \frac{1}{1+\beta^2} = \frac{1}{1+5^2} = \frac{1}{26} = 0.038$

so $\boxed{\alpha \simeq 0.04}$

b) $F_\beta$ is a score that indicates how much more important recall is than precision.

$\beta$ is the parameter that controls a balance between P and R. ie Precision and Recall.

- When $\beta = 1$, $F_1$ is equivalent to HM of P and R.
- When $\beta > 1$, $F_1$ becomes more recall oriented.

__/__/____

- when $\beta < 1$, it becomes more precision oriented.

Here our $\beta = 5$, so our recall is 5 times more important as precision.

Proof: $\alpha = \frac{1}{1+\beta^2}$ so $\alpha$ is inversely proportional to $\beta$.

$\alpha \underset{prop}{\propto} \frac{1}{\beta}$. In the eq$^n$ (1) which is $F = \frac{1}{\frac{\alpha}{P} + \frac{(1-\alpha)}{R}}$

we see that $\alpha \underset{prop}{\propto} \frac{1}{P}$ ie $\alpha \downarrow$ es $P \uparrow$ & overall F will give less weigthage to P

and if $\alpha \downarrow$es then $(1-\alpha) \uparrow$ and F will give more weightage to R.

& hence when $\alpha \downarrow$ or $\beta \uparrow$, recall is emphasized.

__/__/__

Q4.) To prove: Gamma distribution belongs to same family of curve as poisson distribution.

We'll prove this in two points -

(i) Gamma distribution belongs to exponential family

(ii) Poisson's distribution belongs to exponential family.

A distribution is in exponential family if we can express the distribution in form $F(x) = \exp\left(\frac{\theta x - b(\theta)}{a(\eta)} + c(x,\eta)\right)$

(i) Gamma distribution

$$F(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \qquad x, \alpha, \beta > 0$$

log both sides

$$\log F(x) = \alpha \log(\beta) - \log(\Gamma(\alpha)) + (\alpha-1)\log(x) - \beta x$$

Now take exp both sides -

$$F(x) = \exp\left(\alpha \log(\beta) - \log(\Gamma(\alpha)) + (\alpha-1)\log(x) - \beta x\right)$$

$$F(x) = \exp\left(-\beta x + \alpha \log \beta + (\alpha-1)\log x - \log(\Gamma(\alpha))\right)$$

~~F(x) =~~

$$F(x) = \exp\left(\frac{\left(\frac{-\beta x}{\alpha} + \frac{\alpha \log \beta}{\alpha}\right)(-1)}{\frac{1/\alpha}{(-1)}} + (\alpha-1)\log x - \log(\Gamma(\alpha))\right)$$

$$F(x) = \exp\left(\frac{\frac{\beta}{\alpha}x - \log\beta}{-1/\alpha} + (\alpha-1)\log x - \log(\Gamma(\alpha))\right) \quad - (1)$$

$$\theta = \frac{\beta}{\alpha}, \quad \eta = \frac{1}{\alpha}, \quad a(\eta) = \frac{-1}{\alpha}.$$

$$\beta = \theta\alpha = \frac{\theta}{\eta} \Rightarrow \log\beta = \log\theta - \log\eta$$

Substituting these in (1), we get

$$= \cancel{\exp\left(\frac{\theta x - \log\theta}{-\eta} + c(x,\eta)\right)}$$

$$\cancel{b(\theta) = \log\theta}$$

$$\cancel{a(\eta) = -\eta}$$

$$= \exp\left(\frac{\theta x - (\log\theta - \log n)}{-n} + \left(\frac{1}{n}-1\right)\log x - \log\left(\Gamma\left(\frac{1}{2}\right)\right)\right)$$

$$= \exp\left(\frac{\theta x - \log\theta}{-n} - \frac{\log n}{n} + \left(\frac{1}{n}-1\right)\log x - \log\left(\Gamma\left(\frac{1}{n}\right)\right)\right) \quad - ②$$

Let $c(x,n) = \left(\frac{1}{n}-1\right)\log x - \frac{\log n}{n} - \log\left(\Gamma\left(\frac{1}{n}\right)\right)$

→ a function of $x$ & $n$.

substituting these ②, we get

$$= \exp\left(\frac{\theta x - \log\theta}{-n} + c(x,n)\right)$$

$b(\theta) = \log\theta$. $\quad a(n) = -n$

$$\boxed{F(x) = \exp\left(\frac{\theta x - b(\theta)}{a(n)} - c(x,n)\right)}$$ → equation of exponential family

∴ Gamma distribution belongs to exponential family.

(ii) Poisson's Distribution

$$\boxed{F(x) = \frac{\lambda^x e^{-\lambda}}{x!}} \qquad \lambda > 0, \; x = 0, 1, 2 \ldots$$

Take log on both sides –

$\log(F(x)) = x\log\lambda - \lambda - \log(x!)$

take exp both sides –

$F(x) = \exp(x\log\lambda - \lambda - \log(x!)) \quad - ①$

$\theta = \log\lambda$, $\lambda = e^\theta$

$b(\theta) = e^\theta$, $a(n) = 1$, $c(x,n) = -\log(x!)$

substituting these values in ①, we get

$F(x) = \exp(\theta x - e^\theta + c(x,n))$

$$\boxed{F(x) = \exp\left(\frac{\theta x - b(\theta) + c(x,n)}{a(n)}\right)}$$

∴ Poisson Distribution belong to exponential family.

From (i) & (ii) we can say poisson & gamma belongs to same family of curves.

/ /

Q5) Lets assume / consider that Nilesh computes co-variance at every iteration. He is using covariance with k-means & obtaining clustures using co-variance.

(i) k-means algorithm clusters the points with respect to distance If the distance of a point from lets say clusture $C_1$ is more than cluster $C_2$, then k-means places that point to clusture $C_2$. But in this approach, there is a possibility that co-variance indicates that point belong to $C_1$. so this approach may place point in wrong clustures.

(ii) As k-means uses distance & follow hard clustering & places the point to the clusture having minimum distance But covariance is a kind of probabilistic to explain how points are related. so combinely, this comb$^n$ of k-means & co-variance may not perform well.

Scanned with CamScanner