Ayush Aggarwal MT22095

Q3(a)

(i) In one-vs-all classification, we use binary classifier n times; each time over entire dataset using which we identify out one class label over all others. This is combined n times to correctly classify all n labels.

So, 'n' binary classifiers required in one-vs-all.

(ii) In one-vs-one classification, we split the dataset for each pair of classes only & apply the binary classifier model over it. So $^nC_2 = \frac{n(n-1)}{2}$ pairs are formed and hence each classifier

classifies one class opposite to just 1 class. Combining all such solutions, we have classification over entire dataset for every class opposite to every class.

So, $\frac{n(n-1)}{2}$ binary classifiers required in one-vs-one.

Q4.

To prove: gamma distribution belongs to the same family of curves as poisson distribution.

Proof: We shall prove the claim by proving that both gamma & poisson distribution belongs to the same exponential family of distributions.

We prove this in two parts:

(i) gamma distribution belongs to exponential family

(ii) Poisson distribution belongs to exponential family

We know that a distribution is in exponential family, if we can express distribution in following form :-

$$\boxed{P(y;\vec{\eta}) = b(y)\exp\left(\vec{\eta}^{\,}T(y) - a(\vec{\eta})\right)}$$

where,

$y \rightarrow$ data

$\vec{\eta} \rightarrow$ vector of natural parameters

$b(y) \rightarrow$ base measure

$T(y) \rightarrow$ Sufficient statistic

$a(\vec{\eta}) \rightarrow$ log partition

i) Gamma distribution

$$\boxed{P(y;\beta,\alpha) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} y^{\alpha-1} e^{-\beta y}} \qquad y, \alpha, \beta > 0$$

Taking log both sides,

$$\log P(y;\beta,\alpha) = \alpha\log(\beta) - \log(\Gamma(\alpha)) + (\alpha-1)\log(y) - \beta y$$

take exp both sides, getting back the same eqn:

$$P(y;\beta,\alpha) = \exp\left((\alpha-1)\log(y) - \beta y + \alpha\log\beta - \log(\Gamma(\alpha))\right)$$

$$= \exp\left(\left[(\alpha-1), -\beta\right]\begin{bmatrix}\log(y)\\ y\end{bmatrix} - \log\left(\Gamma(\alpha) - \alpha\log\beta\right)\right) \quad (1)$$

$$\vec{\eta} \Rightarrow \left[(\alpha-1, -\beta)\right]^T, \quad T(y) = \begin{bmatrix}\log(y)\\ y\end{bmatrix}, \quad b(y) = 1$$

$$\alpha-1 \downarrow \eta_1, \quad -\beta \downarrow \eta_2$$

Substituting these values in eqn (1) we get

$$= b(y)\cdot\exp\left(\vec{\eta}\,T(y) - \underbrace{\left(\log(\Gamma(\eta_1+1)) - (\eta_1+1)\log(-\eta_2)\right)}_{a(\eta_1,\eta_2)}\right)$$

$$a(\vec{\eta}) = a(\eta_1,\eta_2)$$

$$\boxed{P(y;\eta) = b(y)\cdot\exp\left(\vec{\eta}\,T(y) - a(\vec{\eta})\right)}$$

$\therefore$ gamma Distribution belongs to exponential family

ii) Poisson Distribution :

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

Taking log both sides :

$$\log P(y;\lambda) = y\log\lambda - \lambda - \log(y!)$$

Taking exp both sides to get back same eqn

$$P(y;\lambda) = \exp\left(y\log\lambda - \lambda - \log(y!)\right)$$

$$= \exp\left(-\log(y!)\right)\exp\left(y\log\lambda - \lambda\right)$$

let $b(y) = \exp(-\log(y!))$, $T(y) = y$, $\eta = \log\lambda$ $\quad \lambda = e^{\eta}$

Substituting these values in above equation, we get

$$P(y;\eta) = b(y)\cdot\exp\left(\eta T(y) - e^{\eta}\right)$$

let $a(\eta) = e^{\eta}$

Substituting this, we get

$$\boxed{P(y;\eta) = b(y)\exp\left(\eta T(y) - a(\eta)\right)}$$

$\therefore$ Poisson Distribution also belongs to exponential family, & From (i) & (ii) we can say Gamma Distribution & Poisson Distribution belongs to same family of curves.

Q(7)(a)

Derive : F5 score in terms of precision & recall.
This is to be derived in concurrence with F1 score.

Fβ- score is an adjustable single-score metric used in ML for evaluating binary classification model using the precision & recall values for the class.

The generalised formula for Fβ- score is:

$$F\beta\text{- score} = \frac{(1+\beta^2)(\text{precision} * \text{recall})}{\beta\ \text{precision} + \text{recall}}$$

For F5- score, we will have $\beta = 5$.

$$\text{Hence F5 score} = \frac{(1+5^2)(\text{precision} * \text{recall})}{5^2\ \text{precision} + \text{recall}}$$

$$\boxed{\text{F5 score} = \frac{26(\text{precision} * \text{recall})}{25\ \text{precision} + \text{recall}}}$$

Now, for value of α, we know that

$$\alpha = \frac{1}{1+\beta^2} \quad (\because \beta = 5)$$

$$= \frac{1}{1+5^2} = \frac{1}{26}.$$

$$\boxed{\alpha = 0.0385}$$

Q7(b):

Precision:

① Precision deals with correctly and incorrectly cases predicted as positive.
② Precision is a measure of how many of the positive predictions made are correct (True Positive)
③ We try to minimize false positives in precision.
④ Formula :-

$$\boxed{\text{Precision } P = \frac{TP}{TP + FP}}$$

TP – True Positive
FP – False Positive

Recall:

① Recall is a measure of how many Positive cases the classifier correctly predicted over all the positive cases in data

② In recall, we try to minimize false negatives
③ Formula:

$$\boxed{\text{Recall} = \frac{TP}{TP + FN}}$$

FN → False negative

Ex: Spam email classification

		Predicted class N	Predicted class P
Actual class	N	TN	FP
	P	FN	TP

(Actual)	Predicted Spam	Predicted Ham
Spam	12	14
Ham	0	114

$$\text{Precision} = \frac{14}{14 + 114} = 0.89$$

$$\text{Recall} = \frac{114}{0 + 114} = 1$$

Observations :

1. Decreasing FN increases FP, i.e. when we increase recall, precision decreases
2. When we decrease FP then FN increase, i.e., when we increase precision, recall decreases.
3. In $F\beta$ score, β determines the proportional wt. given to precision over recall in calculation of $F\beta$ score.

Having high β means $F\beta$ score will be sensitive for a given precision and could $\downarrow$ a lot for slight $\uparrow$ in Precision. Hence, precision must be low & recall high for good $F\beta$ score.