Q1. (a) Normalization is used when the data does not have gaussian/Normal distribution where as standardization is used on data having Gaussian/Normal distribution. Normalization is considered when the algorithm do not make assumptions about data distribution while as standardization is used when algorithms make assumptions about the data distribution.

Normalization scales in a range of [0,1] or [-1,1) while as standardization is not bound by range. Normalization is more suitable for this dataset because data does not have a guassian distribution.

Q3. (a)

- One vs all → N binary classifier models
- One vs one → $\frac{N(N-1)}{2}$ binary classifier models

Q4. (a)

Gamma distribution:

$$\text{pdf}: \quad f(y:\alpha,\beta) = \frac{y^{\alpha-1}\beta^{\alpha}e^{-\beta y}}{F(\alpha)} \quad \text{—(1)}$$

We have to prove this is a part of same family of curves as poisson distribution i.e., exponential family.

∴ we need to write it as:

$$\exp\left(\frac{\theta y - b\theta}{a(\phi)} + c(y:\phi)\right)$$

constant term

Taking log on both sides of $=n$ ①:

$$\Rightarrow \log f(y:\alpha,\beta) = \log(y^{\alpha-1}\beta^{\alpha}e^{-\beta y}) - \log F(\alpha)$$

$$\Rightarrow \log f = (\alpha-1)\log y + \alpha\log\beta - \beta y - \log F(\alpha)$$

now, in exponential form:

$$\Rightarrow f = \exp\{(\alpha-1)\log y + \alpha\log\beta - \beta y - \log F(\alpha)\}$$

$$= \exp\{-\beta y + \alpha\log\beta + (\alpha-1)\log y - \log F(\alpha)\}$$

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

dividing & multiplying this term by α

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

Now,

let $\theta = \beta/\alpha$, $a(\phi) = -1/\alpha$, $\phi = 1/\alpha$

$\Rightarrow \boxed{\beta = \theta\alpha}$ also,

$\Rightarrow \boxed{\beta = \frac{\theta}{\phi}}$ $\boxed{\log\beta = \log\theta - \log\phi}$

$$\therefore f(y:\alpha,\beta) = \exp\left\{\frac{y\theta - \log\theta}{-\phi} + \underbrace{\frac{\log\phi}{\phi} + \left(\frac{1}{\phi}-1\right)\log y - \log F(1/\phi)}_{\substack{\text{constant} \\ c(y:\phi)}}\right\}$$

$$\boxed{\therefore f(y:\alpha,\beta) = \exp\left\{\frac{y\theta - \log\theta}{-\phi} + c(y:\phi)\right\}}$$

gamma distribution as exponential

therefore gamma distribution belongs to same distribution of curves as poisson distribution i.e., exponential family.

Now, mean $(b'(\theta)) = \frac{\partial}{\partial\theta}\log\theta = 1/\theta$

canonical link $[b']^{-1} = b(\theta) = \log\theta$

$\therefore (b')^{-1} = 1/\mu$

in terms of μ

variance function $b''(\theta) = 1/\phi^2$

Variance : $a(\phi)\, b''(\theta) = P/\phi^2$.

Q5. (a) The two reasons that the professor pointed out for nilesh's algorithm are as follows:

(i) He calculated covariance matrix for each group which would only gauge the direction of the relationship. However, it does not indicate the strength of the relationship nor the dependency between the variables. So correlation matrix would be better in this case.

~~(ii) He should have calculated covariances~~

(ii) Also covariance is affected by the change in scale. If all the values of one variable are multiplied by a constant and all the values of another variable are multiplied by a similar or different constant than the covariance is changed.

Q6. (c) No, this approach is not at all greedy, because we do not solve a problem by selecting the best option available at the moment as we try to find the best option by trying to eliminate the weaker features and see the accuracy after each elimination. Here, we try to get accuracy of the model after every feature selection & extraction methods.

Q7. (a)

The full definition of the F-measure is as follows:

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

where,

$\beta \rightarrow$ controlls balance b/w P & R $(0 \leq \beta \leq \infty)$

$P \rightarrow$ precision, $R \rightarrow$ recall

Also, we know that

$$\alpha = \frac{1}{\beta^2+1} \Rightarrow \boxed{\alpha = \frac{1}{5^2+1} = \frac{1}{26}}$$

Now, for F_5 score $\rightarrow \beta = 5$

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

(b) As $(\beta > 1)$ F scores becomes more recall oriented i.e., recall is more emphasised here. We know that recall represents the models ability to correctly predict the positive out of actual positives. Recall is also known as sensitivity or true positive rate.

This score is good for real-world problems like cancer or covid detection as more the emphasis on recall less the chance of not detecting cancer or covid.