

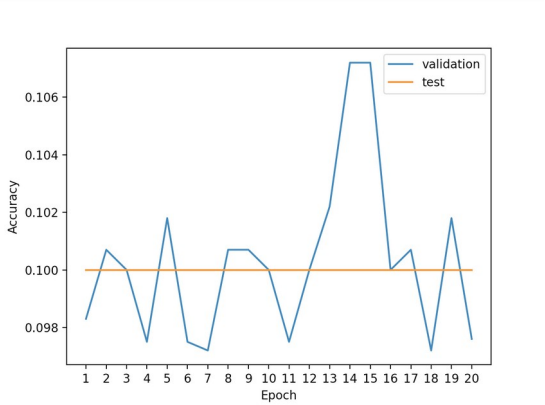
H1  
GROUP 51  
João Moniz nº 83480  
Liana Baghdasaryan nº 104708

Work Distribution  
Liana Baghdasaryan: q1  
João Moniz: q2 and q3

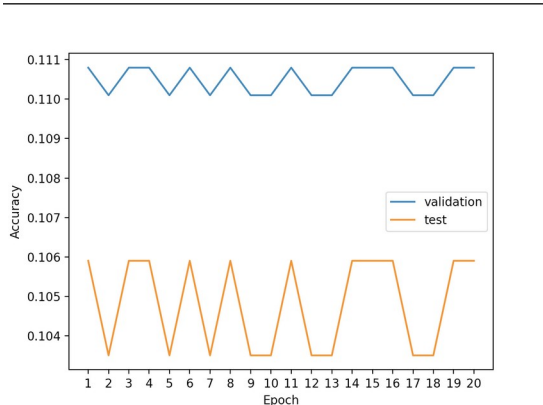
Question 1:

1.1

a)

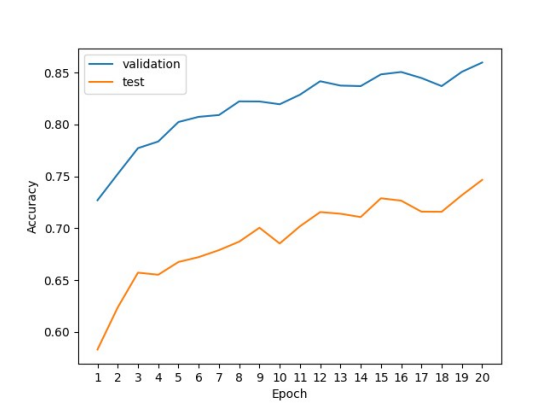


b)



1.2.

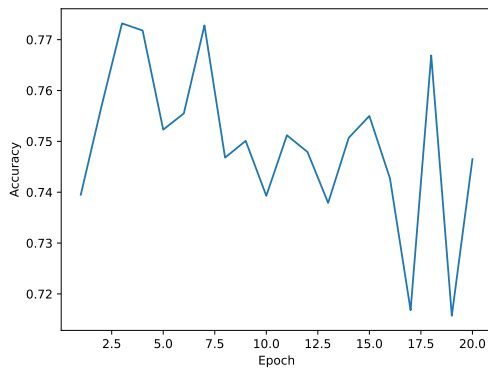
b)



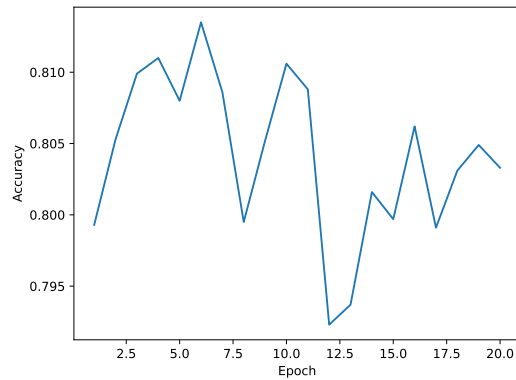
## Question 2:

### 2.1:

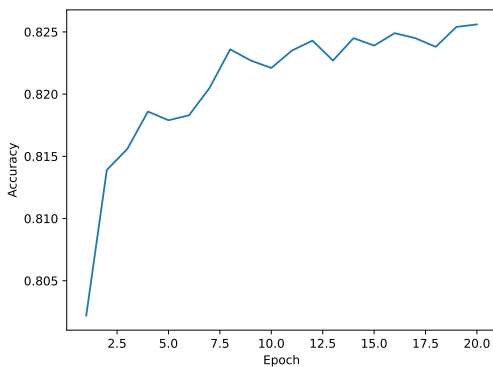
Learning Rate 0.1  
(F.T.A.=0.6151)



L.R. = 0.01  
(F.T.A.=0,6806)



L.R.=0.001  
(Final Validation Accuracy 0,7019)



Best learning rate result was 0.001 with 0,7019 F.V.A..

### 2.2:

best

number of epoch = default

learning rate = default

hidden layer size = 200

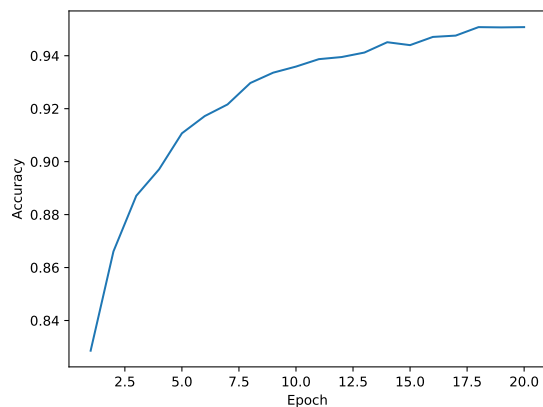
drop out = default

batch size = default

activation = default

optimizer = default

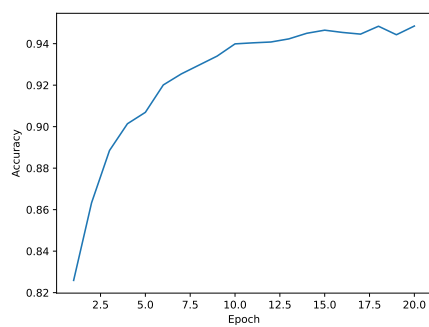
F.V.A =0,8819



2.3:

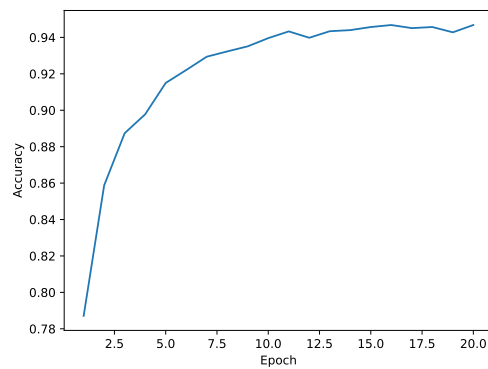
n° of hidden layers= 2

F.V.A.= 0,8815



n° of hidden layers=3

F.V.A. = 0,8764



Best result is model with 2 hidden layers.

Question 3:

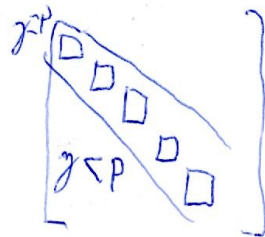
1 and 2

q) 3.1) If  $h$  is a linear transformation of  $\phi(x)$  then

$$h = A_\theta \phi(x)$$

$$h_i = \left( \sum_{j=1}^D w_{ij} x_j \right) \times \left( \sum_{p=1}^D w_{ip} x_p \right)$$

$$= \sum_{j,p} (w_{ij} w_{ip}) (x_j x_p)$$



$$= \sum_{j=p} w_{ij} \cdot w_{ip} \cdot x_j \cdot x_p + 2 \sum_{j < p} w_{ij} \cdot w_{ip} \cdot x_j \cdot x_p$$

$$= \sum_{j=p} w_{ij} \cdot w_{ip} \cdot x_j \cdot x_p + \sum_{j < p} 2 w_{ij} \cdot w_{ip} \cdot x_j \cdot x_p$$

q) 3.1)  $\phi(x)$   $\rightarrow$   $\phi_j, p$   
 $q = (j, p)$   
 $j < p$

$$= \sum_{q=1}^{D(D+1)/2} (A_\theta)_{iq} (\phi(x))_q$$

$$R: (A_\theta)_{iq} = \begin{cases} w_{ij} w_{ip} & \text{if } j=p \\ 2 w_{ij} w_{ip} & \text{if } j < p \end{cases}$$

$$(\phi(x))_q = x_j x_p$$

q) 3.2)

$$\hat{y} = V^T h$$

or seen in 3.2)

$$\hat{y} = V^T A_\theta \phi(x)$$

$$= [V]^T [A_\theta] [\phi(x)]$$

$$\hat{C}_\theta = V^T C_\theta$$

$$C_\theta = A_\theta^T V$$

$$C_\theta^T = V^T A_\theta$$

R: No, because there are still the necessity of  
 parameter changes