1. **Explain how the "pixel similarity" approach to classifying digits works.**

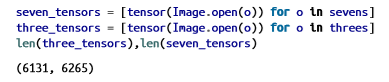
Der laves et tensor stack, hvor der tages et average af hvert pixel, som så sammenlignes med det ideelle billed af et 3- eller 7- tal

1. **What is a list comprehension?**

Syntax: list= [*expression* for item in iterable if condition]

En list comprehension er en kort måde at lave en liste på.

Eksempler:

  
Create one now that selects odd numbers from a list and doubles them.

double\_odd = [double\_numbers(i) for i in numbers if i%2 == 1]

1. **What is a "rank-3 tensor"?**

Det er en liste af rank-2 matrix

1. **What are RMSE and L1 norm?**

Root Mean Square Error : forskellene mellem værdier forudsagt af en model eller en estimator og de observerede værdier.

L1 norm : The L1 norm that is calculated as the sum of the absolute values of the vector.

1. **Create a 3×3 tensor or array containing the numbers from 1 to 9.   
   Double it.**

  
Select the bottom-right four numbers.

1. **What is broadcasting?**

Det er sort magi der udvider tensor med end mindre rank til at matche den med en stør rank.

1. **Are metrics generally calculated using the training set, or the validation set?**

Validation

1. **What is SGD?**

Stochastic Gradient Descent :

Det er en optimizer der kan håndtere stor datasets idet den samler det i mini-batches

1. **What are the seven steps in SGD for machine learning?**
   1. Init
   2. predict
   3. loss
   4. gradient
   5. step
   6. repeat?
   7. stop
2. **How do we initialize the weights in a model?**

vi tager random værdier for at starte med, da det alligevel bliver tilpasset

1. **What is "loss"?**

loss is a number indicating how bad the model's prediction was on a single epoch. If model prediction is perfect the loss is 0.

1. **Why can't we always use a high learning rate?**

Med en for stor learning rate, laver vi for store steps, hvormed vi ikke kan komme til bedste værdier. Den skyder over målet

1. **What is a "gradient"?**

En gradient er værdien på et fald eller stigning fra et punkt på en graf.

“Gradient is denied as rise/run; that is, the change in the value of the function, divided by the change in the value of the parameter.”

1. **Why can't we use accuracy as a loss function?**

accuracy er for mennesker

loss bliver brugt for modellen

1. **What is the difference between a loss function and a metric?**

loss er antallet af forkerte predictions, mens metric er optimeret for mennesker, det viser i % hvor højt scoren er i forhold til rigtige predictions

1. **What is the function to calculate new weights using a learning rate?**

w -= gradient(w) \* lr

1. **What does the backward method do?**

backwards methoden laver backpropagation, som optimizer weights, og dermed “træner” modellen