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

* Finder gennemsnitlig værdi for hver individuelle pixel i billedet og laver et ideelt billede. Derefte sammenligner den med de ideelle billeder for at karakterisere tallet.

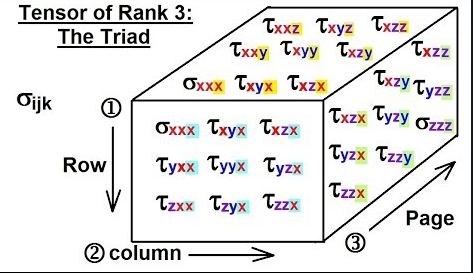
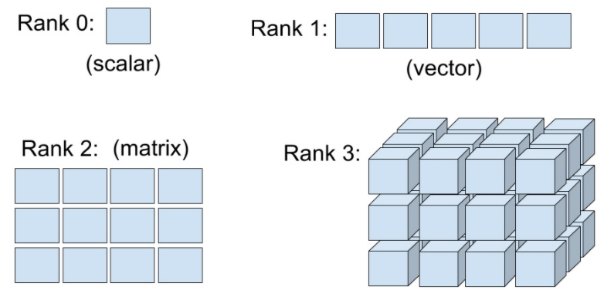
**2. What is a list comprehension? Create one now that selects odd numbers from a list and doubles them.**

* Måde at lave en liste (LINQ)

|  |
| --- |
| numbers = [1,2,3,4,5,6,7,8,9] odd = [x\*2 for x in numbers if x%2==0] |

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

* Tre dimensionel tensor



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

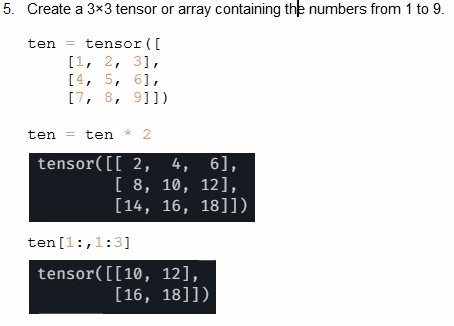
* **L1 norm:**
* Take the mean of the *absolute value* of differences (absolute value is the function that replaces negative values with positive values). This is called the *mean absolute difference* or *L1 norm*

* **(L2 norm) RMSE:**
* Take the mean of the *square* of differences (which makes everything positive) and then take the *square root* (which undoes the squaring). This is called the *root mean squared error* (RMSE) or *L2 norm*.

**5. Create a 3×3 tensor or array containing the numbers from 1 to 9. Double it. Select the bottom-right four numbers.**

|  |
| --- |
| data =  [[1,2,3],[4,5,6],[7,8,9]]  tens = tensor(data) \* 2  tens[-2:, -2:] |

eller:



**6. What is broadcasting?**

* Broadcasting means it will automatically expand the tensor with the smaller rank to have the same size as the one with the larger rank.

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

* Metrics are always calculated on the validation set.
* Fordi Metrics fortæller os om kvaliteten af vores model

**8. What is SGD?**

* Stochastic gradient descent (often abbreviated SGD) is an [iterative method](https://en.wikipedia.org/wiki/Iterative_method) for [optimizing](https://en.wikipedia.org/wiki/Mathematical_optimization) an [objective function](https://en.wikipedia.org/wiki/Objective_function) with suitable [smoothness](https://en.wikipedia.org/wiki/Smoothness) properties (e.g. [differentiable](https://en.wikipedia.org/wiki/Differentiable_function) or [subdifferentiable](https://en.wikipedia.org/wiki/Subgradient_method)).

**9. What are the seven steps in SGD for machine learning?**

* Initialize the weights.
* For each image, use these weights to predict whether it appears to be a 3 or a 7.
* Based on these predictions, calculate how good the model is (its loss).
* Calculate the gradient, which measures for each weight, how changing that weight would change the loss
* Step (that is, change) all the weights based on that calculation.
* Go back to the step 2, and repeat the process.
* Iterate until you decide to stop the training process (for instance, because the model is good enough or you don't want to wait any longer).

**10. How do we initialize the weights in a model?**

* Random values

**11. What is "loss"?**

* A measure of how good the model is, chosen to drive training via SGD(Stochastic Gradient Descent)

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

* Risikere at “hoppe” (step over) vores optimal/local minimum
* Kan risikere at få en forværret loss
* Risiko for at den “bouncer” som kan resultere i at der skal betydeligt flere trainings til at få trænet med succes

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

* En gradient måler ændringen i alle vægtene i forhold til ændringen i error.
* Gradient is simply a vector which gives the direction of maximum rate of change. By taking steps in that direction, we hope to reach our optimal solution.

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

Gradient can be written in this in mathematically as: (y\_new - y\_old) / (x\_new - x\_old). This gives us a good approximation of the gradient when x\_new is very similar to x\_old, meaning that their difference is very small.   
But accuracy only changes at all when a prediction changes from a 3 to a 7, or vice versa. The problem is that a small change in weights from x\_old to x\_new isn't likely to cause any prediction to change, so (y\_new - y\_old) will almost always be 0. In other words, the gradient is 0 almost everywhere.

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

Loss: A measure of how good the model is, chosen to drive training via SGD(Stochastic Gradient Descent)

Metric: A measurement of how good the model is, using the validation set, chosen for human consumption

* A loss function is used to **train** your model. A metric is used to **evaluate** your model.
* A loss function is used **during** the learning process. A metric is used **after** the learning process

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

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