Assignment 1

1) How does the choice of the number of neighbors K affect the performance of KNN Classifier?

- A smaller K may result in a more flexible model, sensitive to noise
- A larger K may lead to a smoother decision boundary but could overlook local patterns

2) What are the advantages and disadvantages of using KNN for image classification?

advantages	disadvantages
- simplicity	- sensitive to noise
- no training phase	- Computational Cost
- effectiveness for simple patterns	- High dimensional data
	- Memory requirement

Assignment 2

1) What happens to softmax loss when the predicted probability for the true label increase?

- The softmax loss will decreases

2) What is the key difference between softmax loss and multi class SVM loss in terms of their approch to handling multi class classification?

- softmax loss uses probabilities and penalizes confidently incorrect predictions
- SVM loss directly operates on raw scores and penalizes the margin between the correct class score and other class scores

3) For the multiclass SVM loss $L_i = \sum_{j \neq y_i} \max(0, s_j - s_{y_i} + 1)$

(a) At initialization W is small so all $s \approx 0$. what is the loss L_i assuming N examples and C classes?

$L_i = C - 1$ because the loss is $\max(0, 0 - 0 + 1) = 1$

(b) What if the sum was over all classes ? including $j = y_i$

$L_i = C$

4) For the following scores calculate softmax and multi SVM loss if the first class is the correct class

a. $[10, -100, -100]$ b. $[10, 9, 9]$

What does the calculated loss tells you about the difference between the two losses?

(a) $\text{softmax loss} = -\log\left(\frac{e^{10}}{e^{10} + e^{-100} + e^{-100}}\right) = 0$

$\text{SVM loss} = \max(0, -100 - 10 + 1) + \max(0, -100 - 10 + 1) = 0$

(b) $\text{softmax loss} = -\log\left(\frac{e^{10}}{e^{10} + e^{9} + e^{9}}\right) = 0.5514$

$\text{SVM loss} = \max(0, 9 - 10 + 1) + \max(0, 9 - 10 + 1) = 0$

The calculated tell you that softmax loss is more sensitive to the difference between the correct score and the others scores but SVM loss only cares that the correct score is higher than the others by some margin

Assignment 3

1) what is the purpose of gradient descent optimization in machine learning?

- minimize the loss function or cost by computes gradients and updates parameters

2) What is the role of batch size in gradient descent optimization?

- represents the number of training examples used in each iteration it influences the update frequency of the model parameters

3) What is the learning rate in gradient descent and why is it important?

- learning rate is a hyperparameter the determines the stepsize at each iteration
- important → it affects the convergence and performance of gradient descent

4) Consider a quadratic function $f(X) = X^2 - 4X + 3$. use gradient descent to find minimum value of the function. start with an initial guess for X and update it iteratively using the gradient descent algorithm use the following conditions

initial guess $X = 0$ learning rate : 0.25

Num of iterations : 5

gradient $\nabla f(X) = 2X - 4$ update $X = X - \text{learning rate} * \nabla f(X)$

→ iteration 1

$X = 0$

$\nabla f(X) = 2*0 - 4 = -4$

update $X = 0 - 0.25 * (-4) = 1$

→ iteration 2

$X = 1$

$\nabla f(X) = 2*1 - 4 = -2$

update $X = 1 - 0.25 * (-2) = 1.5$

→ iteration 3

$X = 1.5$

$\nabla f(X) = 2*1.5 - 4 = -1$

update $X = 1.5 - 0.25 * (-1) = 1.75$

→ iteration 4

$X = 1.75$

$\nabla f(X) = 2*1.75 - 4 = -0.5$

update $X = 1.75 - 0.25 * (-0.5)$

$= 1.875$

→ iteration 5

$X = 1.875$

$\nabla f(X) = 2*1.875 - 4 = 0.75$

update $X = 1.875 - 0.25 * 0.75$

$= 1.6875$

$f(1.6875) = X^2 - 4X + 3$

$= 0.9219$

Assignment 4

1) what is the primary difference between traditional Computer program and neural network in terms of problem solving ?
- Traditional Computer program → follow explicit instructions and rules
- Neural networks → learn patterns and relationships from data and make them more adept at handling complex tasks

2) You have a neural network with 3 layers : an input layer with 100 neurons a hidden layer with 50 neurons and an output layer with 10 neurons - if each neuron in the hidden layer is connected to each neuron in the input and output layers. Calculate the num of weights in the network ?
- Weights = 100 * 50 + 50 * 10 = 5500

3) suppose you are training a neural network for image classification with a batch size of 32, and you have 10000 training images. How many weight updates will occur during one epoch (one pass thrugh the entire training dataset) ?
- num of weight update = 10000 / 32 = 312.5 ≈ 313

4) what is backpropagation and why is it important in training neural networks ?
- backpropagation → recursive application of chain rule along computational graph to compute the gradient of all inputs / parameters / intermediates
- important → adjusting network's parameters to minimize the difference between predicted and actual outputs

5) Given an image with dimensions 100*100 pixels and Convolutional layer with a 3*3 filter and no padding, Calculate the dimensions of feature map produced?

- outsize = (inputsize - filtersize + 2*padding) / stride + 1
  inputsize=100 filtersize=3 padding=0
  let stride=1
  outsize = (100 - 3 + 2*0)/1 + 1 = 98
  resulting in a 98*98 feature map

6) What are filters (Kernels) in a Convolutional Layer, and how are they used to extract features from images?

- filters is a Convolutional layer are small windows used to extract features from input data such as images
- by learning filters, Convolutional layers can automatically detect patterns like edges, textures or more complex structures, enabling the network to learn hierarchical representations of the input data

---

Assignment 5

1) What is a Convolutional Neural Network (CNN) and what makes it suitable for image-related tasks?

- CNN → type of neural network designed for processing structured grid data such as image
- it ability to automatically learn hierarchical representations, capturing local patterns and global structures

2) Explain the Concept of local Connectivity in CNNs?

- refers to the concept of neurons being connected to a small, localized

region of the input volume rather than entire input

3) What are the Key parameters of a Convolutional layer?

- num of filters K
- filter size F
- stride s
- zero padding p

4) Why is zero padding used in Convolutional layer, and what impact does it have on the spatial arrangement?

- zero padding is used in Convolutional layer to maintain the spatial dimensions of the input volume
- Zero padding has several impacts on spatial arrangement:-
  1- preserving spatial size
  2- Centering receptive field
  3- Border information
  4- Mitigating information loss

5) What is the purpose of pooling layer in ConvNet?

- makes the representations smaller and more manageable
- operates over each activation map independently

6) if a pooling layer with a filter size 2x2 and stride of 2 is applied to an input volume of size 28 x 28 x 64, Calculate the size of output volume

Applying pooling layer → the output volume will be 14 x 14 x 64

7) Given an input volume of size 32 x 32 x 3 and Convolutional layer with 32 filter of size 3x3 and stride of 1, Calculate the size of output volume

output volume = (W-f+2p)/s+1 = (32-3+2x0)/1+1 = 30

30 * 30 * 32

8) For a ConVolutional layer with 8 filter of a size 3 X 3 and a depth of 3 Calculate the total num of parameters including biases

num of Parameters = size of filter * depth + biases

= 3 X 3 X 3 + 1 = 28 Params

⟶ 28 * 8 = 224

9) For a fully-Connected layer Connected to an input Volume of 256 X 256 X 3 how many weights and biases are needed if there are 100 neurons in the fully-Connected layer?

weights = 256 * 256 * 3 * 100 = 19660800

biases = 1 * 100 = 100 biases