Name: Dipen Prajapati
Net ID: dp1435

HW-4

**Problem – 1**

Ans - (a)

Conventional SGD computes the loss gradient over each mini-batch and updates the parameters with a single and global learning rate. It is easy to implement but tends to oscillate in high-curvature regions and converges slowly.

SGD with Momentum adds a velocity term that accumulates the proportion of previous updates, effectively smoothing the optimization trajectory. By combining current gradients with momentum, it is quicker in coherent directions and suppresses oscillations.

AdaGrad normalizes each parameter's learning rate by dividing by the square root of cumulative squared gradients. This serves to make more substantial updates to less frequently updated parameters (good for sparse data) but may cause the effective learning rate to decline too fast in the long term.

RMSprop modifies AdaGrad by utilizing an exponential moving average of squared gradients, rather than the accumulating sum of squared gradients. This modification prevents the learning rates from being decreased too rapidly and makes the progress more stable and long-lasting.

In the last, Adam ("Adaptive Moment Estimation") unites the momentum (first-moment estimates) with the adaptive second moment estimates of RMSprop and adds bias correction. Its fast convergence with properly fine-tuned default parameters has established Adam as the reference optimizer for most deep-learning tasks amongst others.

Ans - (b)

Results

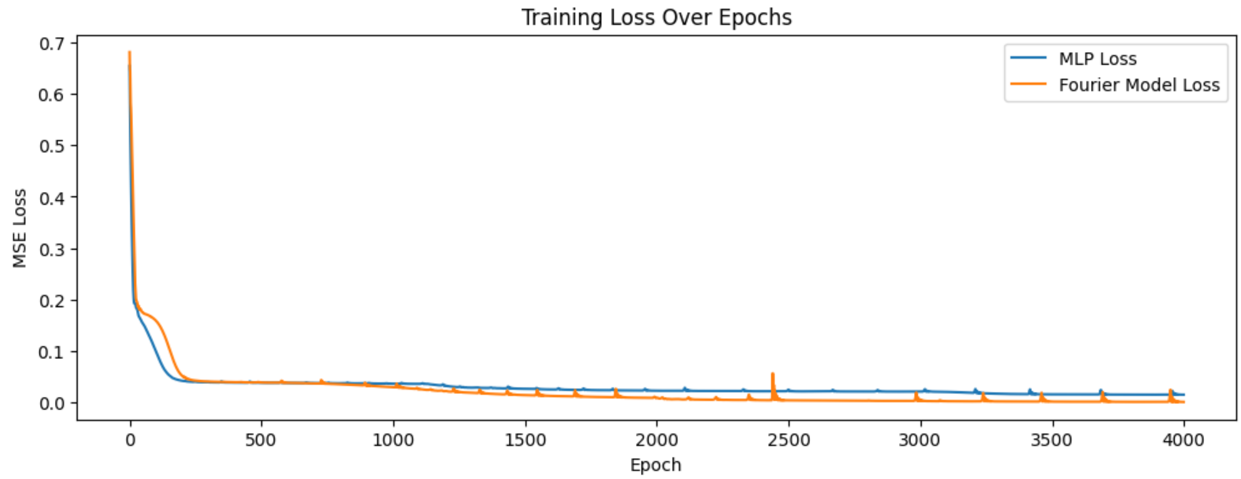
```
Optimizer: SGD
epoch  5 Test Accuracy: 0.9333
epoch 10 Test Accuracy: 0.9422
Final test Acuracy (SGD): 0.9422

Optimizer: SGD_Momentum
epoch  5 Test Accuracy: 0.9755
epoch 10 Test Accuracy: 0.9787
Final test Acuracy (SGD_Momentum): 0.9787

Optimizer: AdaGrad
epoch  5 Test Accuracy: 0.9740
epoch 10 Test Accuracy: 0.9778
Final test Acuracy (AdaGrad): 0.9778
```

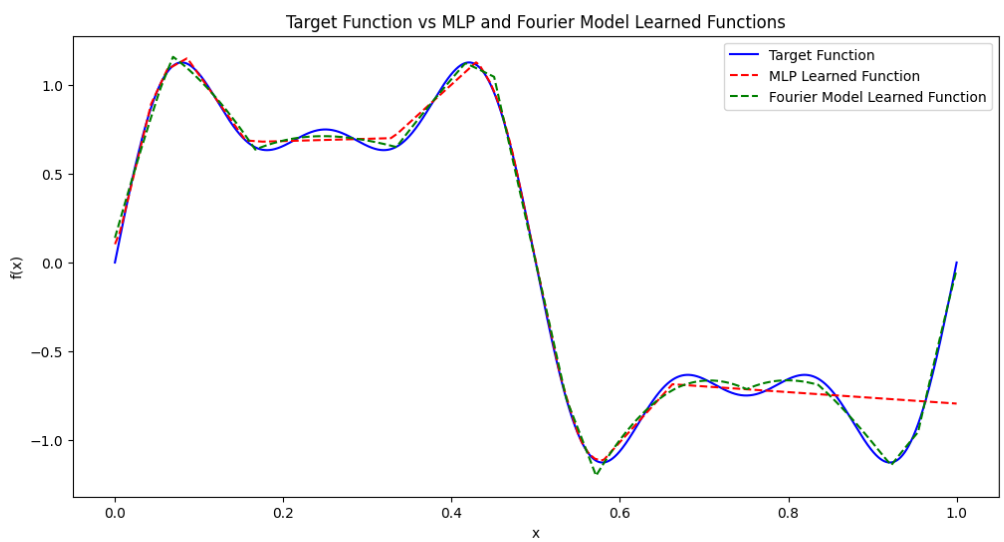
```
Optimizer: RMSprop
epoch  5 Test Accuracy: 0.9744
epoch 10 Test Accuracy: 0.9788
Final test Acuracy (RMSprop): 0.9788

Optimizer: Adam
epoch  5 Test Accuracy: 0.9762
epoch 10 Test Accuracy: 0.9786
Final test Acuracy (Adam): 0.9786

summary of test accuracies:
  SGD            : 0.9422
  SGD_Momentum   : 0.9787
  AdaGrad        : 0.9778
  RMSprop        : 0.9788
  Adam           : 0.9786
```

**<u>Problem – 2</u>**

**Output:**

```
Epoch [100/4000], MLP Loss: 0.1005, Fourier Loss: 0.1595
Epoch [200/4000], MLP Loss: 0.0424, Fourier Loss: 0.0520
Epoch [300/4000], MLP Loss: 0.0397, Fourier Loss: 0.0407
Epoch [400/4000], MLP Loss: 0.0388, Fourier Loss: 0.0395
Epoch [500/4000], MLP Loss: 0.0385, Fourier Loss: 0.0388
Epoch [600/4000], MLP Loss: 0.0381, Fourier Loss: 0.0385
Epoch [700/4000], MLP Loss: 0.0378, Fourier Loss: 0.0372
Epoch [800/4000], MLP Loss: 0.0374, Fourier Loss: 0.0355
Epoch [900/4000], MLP Loss: 0.0371, Fourier Loss: 0.0352
Epoch [1000/4000], MLP Loss: 0.0367, Fourier Loss: 0.0296
Epoch [1100/4000], MLP Loss: 0.0362, Fourier Loss: 0.0263
Epoch [1200/4000], MLP Loss: 0.0322, Fourier Loss: 0.0223
Epoch [1300/4000], MLP Loss: 0.0288, Fourier Loss: 0.0187
Epoch [1400/4000], MLP Loss: 0.0278, Fourier Loss: 0.0163
Epoch [1500/4000], MLP Loss: 0.0266, Fourier Loss: 0.0144
Epoch [1600/4000], MLP Loss: 0.0253, Fourier Loss: 0.0128
Epoch [1700/4000], MLP Loss: 0.0245, Fourier Loss: 0.0116
Epoch [1800/4000], MLP Loss: 0.0239, Fourier Loss: 0.0105
Epoch [1900/4000], MLP Loss: 0.0234, Fourier Loss: 0.0096
Epoch [2000/4000], MLP Loss: 0.0230, Fourier Loss: 0.0089
Epoch [2100/4000], MLP Loss: 0.0229, Fourier Loss: 0.0060
Epoch [2200/4000], MLP Loss: 0.0224, Fourier Loss: 0.0054
Epoch [2300/4000], MLP Loss: 0.0221, Fourier Loss: 0.0049
Epoch [2400/4000], MLP Loss: 0.0219, Fourier Loss: 0.0046
Epoch [2500/4000], MLP Loss: 0.0249, Fourier Loss: 0.0043
Epoch [2600/4000], MLP Loss: 0.0216, Fourier Loss: 0.0039
Epoch [2700/4000], MLP Loss: 0.0215, Fourier Loss: 0.0037
Epoch [2800/4000], MLP Loss: 0.0214, Fourier Loss: 0.0034
Epoch [2900/4000], MLP Loss: 0.0213, Fourier Loss: 0.0031
Epoch [3000/4000], MLP Loss: 0.0212, Fourier Loss: 0.0044
Epoch [3100/4000], MLP Loss: 0.0206, Fourier Loss: 0.0026
Epoch [3200/4000], MLP Loss: 0.0188, Fourier Loss: 0.0023
Epoch [3300/4000], MLP Loss: 0.0170, Fourier Loss: 0.0021
Epoch [3400/4000], MLP Loss: 0.0162, Fourier Loss: 0.0019
Epoch [3500/4000], MLP Loss: 0.0159, Fourier Loss: 0.0020
Epoch [3600/4000], MLP Loss: 0.0157, Fourier Loss: 0.0016
Epoch [3700/4000], MLP Loss: 0.0163, Fourier Loss: 0.0060
Epoch [3800/4000], MLP Loss: 0.0154, Fourier Loss: 0.0013
Epoch [3900/4000], MLP Loss: 0.0153, Fourier Loss: 0.0012
Epoch [4000/4000], MLP Loss: 0.0152, Fourier Loss: 0.0011
```




Ans – 1: Input and Output

The models both take a single scalar input x in the range [0, 1], represented as a [4000 x 1] tensor. The output is also a [4000 x 1] tensor that approximates the target function: $f(x) = \sin(2\pi x) + 0.5 * \sin(6\pi x) + 0.25 * \sin(10\pi x)$. This means each model is learning to map a 1D input to a 1D output that closely follows this multi-frequency sine function.

Ans – 2: role of LLF layer

The Learnable Fourier Feature (LFF) layer calculates a learnable linear transformation of x, scales by π, and then takes the sine. This projects the one-dimensional input into a high-frequency, dense feature space which subsequent layers can then more easily linearly combine to combat the tendency of the vanilla MLP to learn low frequencies first.

Ans – 3: Observations & Connection to UAT

When you run 4 000 epochs, the Fourier-feature model attains significantly lower MSE (approximately 0.001) much faster than the base MLP (0.015) and fits the sharp oscillations exactly, whereas the MLP smooths them. This shows the Universal Approximation Theorem at work: both architectures can theoretically approximate to the target, but the vanilla MLP's spectral bias makes it sluggish on high-frequency detail, while the LFF layer produces those features earlier and achieves far faster, more accurate convergence.