

Q1.

- a. A tensor is a multidimensional matrix containing elements of a single data type. It can be of any shape, as stored in the `tensor.shape` property, and any number of dimensions.
- b. `Torch.mul` computes the element-wise product between tensors, where each element of one tensor is multiplied by the corresponding element of the other. `Torch.matmul` computes the matrix product.
- c. An optimizer implements the entire process for learning new weights. They are defined in `torch.optim`. Three of the optimizers are `torch.optim.Adadelta`, `torch.optim.Adagrad`, and `torch.optim.Adam`.
- d. `loss = torch.nn.MSELoss()`
`loss(output, target):q`
- e. The `__init__` method instantiates the class, specifying the number of layers and number of nodes in each layer.
The forward method specifies the activation function used for each layer.
- f. A frozen parameter won't have its gradient computed during gradient descent. These are useful for fine-tuning pre-trained models. A parameter can be marked as frozen by setting `Tensor.requires_grad` to `False`.

Q3.

| Epoch num | Train loss | Dev loss |
|-----------|------------|----------|
| 0 | 0.54577 | 0.39595 |
| 1 | 0.31534 | 0.33952 |
| 2 | 0.21518 | 0.33031 |
| 3 | 0.13682 | 0.37068 |
| 4 | 0.08616 | 0.43732 |
| 5 | 0.04479 | 0.59443 |

Test set accuracy of best model: 0.8585437979844525

Total runtime: 431 s

Q4.

| Epoch num | Train loss | Dev loss |
|-----------|------------|----------|
| 0 | 0.61634 | 0.55047 |
| 1 | 0.47888 | 0.44735 |
| 2 | 0.41480 | 0.40235 |
| 3 | 0.36737 | 0.36708 |
| 4 | 0.33170 | 0.38633 |
| 5 | 0.30369 | 0.35311 |

Test set accuracy of best model: 0.8433743605528341

Total runtime: 483 s

Q5.

| Epoch num | Train loss | Dev loss |
|-----------|------------|----------|
| 0 | 0.61634 | 0.55047 |
| 1 | 0.47888 | 0.44735 |
| 2 | 0.41480 | 0.40235 |
| 3 | 0.36737 | 0.36708 |
| 4 | 0.33170 | 0.38633 |
| 5 | 0.30369 | 0.35311 |
| 6 | 0.27000 | 0.32964 |
| 7 | 0.25154 | 0.32151 |
| 8 | 0.22056 | 0.37461 |
| 9 | | |
| 10 | | |
| 11 | | |

Test set accuracy of best model: 0.8574648337900791

Total runtime: 749 s

Please waive the late penalty (second time used)