Hyperparameter Report for HW3

Below are some of the hyperparameters that were tuned for the homework:

- Weight Initialization Scheme
- Minibatch Size
- Activation Function
- Learning Rate

Network Architecture:

Input (35 features) -> Linear (35, 256) -> ReLU -> Linear (256, 64) -> ReLU -> Linear (64, N_Classes) -> Softmax Cross Entropy Loss/ Return Class(Bed Count) with highest scoreMax Index (depending on train/test).

N_Classes is dynamically determined, based on the values in training labels.

Minibatch Size

The size of the mini-batch is crucial in ensuring the timely convergence of the model and its generalization power. The following mini-batch sizes were experimented with: 16, 32, 64, 128, 256.

The other hyperparameters were kept fixed:

- ReLU activation function
- learning rate 0.3
- He initialization
- 200 epochs

	Minibatch Size				
Split No.	16	32	64	128	256
Split 1	0.4019	0.4250	0.4475	0.4502	0.4570
Split 2	0.4407	0.4178	0.4521	0.4585	0.4257
Split 3	0.4547	0.4679	0.4533	0.4777	0.4791
Split 4	0.3756	0.4160	0.4850	0.4662	0.4536
Split 5	0.4283	0.4658	0.4393	0.4658	0.4452

Weight Initialization Scheme

It has been shown in the literature that an educated initialization of weights can drastically improve the convergence and performance of a neural net. The following initialization schemes (for Linear Layer weights and biases) were tested:

He initialization, Random Sampling from a Normal Distribution with mean 0, variance 1, and Uniform Distribution of (0,1).

The other hyperparameters were kept fixed:

- ReLU activation function
- learning rate 0.3
- Minibatch size 128
- 200 epochs

	Weight Initialization		
Split No.	He initialization	Normal Sampling	Uniform
Split 1	0.4502	0.3140	0.3245
Split 2	0.4585	0.3092	0.3085
Split 3	0.4777	0.3154	0.3091
Split 4	0.4662	0.3066	0.3080
Split 5	0.4658	0.3144	0.3268

Activation Function

The following activation functions were tested: **ReLU, Sigmoid, Tanh**
The other hyperparameters were kept fixed:

- He initialization
- learning rate 0.3
- Minibatch size 128
- 200 epochs

	Activation Function		
Split No.	ReLU	Sigmoid	Tanh
Split 1	0.4502	0.3562	0.4141
Split 2	0.4585	0.39	0.4385
Split 3	0.4777	0.3586	0.3850
Split 4	0.4662	0.3191	0.4229
Split 5	0.4658	0.3739	0.4077

Learning Rate

The following learning rates were tested: 0.001, 0.01, 0.1, 0.2, 0.3, 0.5
The other hyperparameters were kept fixed:

- He initialization
- ReLU activation function
- Minibatch size 128
- 200 epochs

	Learning Rate					
Split No.	0.001	0.01	0.1	0.2	0.3	0.5
Split 1	0.2520	0.3140	0.4168	0.4475	0.4502	0.4495
Split 2	0.18	0.3107	0.435	0.4114	0.4585	0.4421
Split 3	0.2381	0.3091	0.4526	0.4693	0.4777	0.4714
Split 4	0.2968	0.3135	0.4383	0.4759	0.4662	0.4432
Split 5	0.2064	0.3144	0.4202	0.4467	0.4658	0.4627