Training error vs Test error Network characteristics Training Error Number of Hidden Layers: 1 10.0 Test Error Units per Layer: 32 9.0 Activation function: Leaky ReLU Batch-size: 1 8.0 Loss function: MEE 7.0 Learning Rate: 1e-05 Regularization: elastic 6.0 *α*: 0.5 *λ*: 1e-06 5.0 Optimizer: NAG **MEE Error** Momentum: 0.9 4.0 3.0 2.0 200 400 600 800 1000 **Epochs**