



(1) Cost function \rightarrow sparse categorical

(2) layers \rightarrow Deep/shallow

We can vary Batch size, maybe add regularization
Init weights with a seed

(1) Autoencoder (2) 10 layer w/o Regularization
① ③ 10 layer w Regularization
Relut Softmax

(3) 2 layer, 1 layer
no Ω $\|w\|_2$

⑤ $784 \rightarrow 397 \approx 10$

⑥ $784 \rightarrow 90 = 90 = 10$

[PTO] for arch/1 5 NN ⑦ \leftrightarrow ⑧
Glorot & HE uniform

⑪ tanh, sigmoid, softmax, selu, relu
hard tanh,

momentas
└─┬─┐
└─┬─┐

(13)

Change Optimizer

└─┬─┐
└─┬─┐
Adam
SGD with momentum

(134)

This is a 5 layer DNN; RELU; no regularization

