Assignment 2 writeup

Thomas Lu

1 Problem 1

(c) Placeholders are nodes in a TF computation graph whose values are populated at runtime. These nodes are often used to populate training data. Feed dictionaries, meanwhile, are what actually do the populating at runtime; they specify at runtime a mapping of placeholder nodes to actual values.

(e) When `train_op` is called, we compute during forward propagation (for a batch) predictions $\hat{y} = \text{softmax}(xW + b)$ and the loss $J = CE(y, \hat{y})$, where CE denotes cross-entropy loss. Then we compute during backpropagation the partial gradients $\partial J/\partial W$ and $\partial J/\partial b$ and add a negative multiple of these gradients to our variables W and b.