Question 3:

(e)
For fully connected model:

kl_divergence_ml loss function gets the best result in validation and test data, while softmax_cross_entropy_with_logits loss function gets the minimum train error. Expected_cost loss function gets the worst result this time, its train loss, valid error and test error are all the maximum values among four models.

Training and testing performance are almost monotonically related. Only one exceptional case is that the softmax_cross_entropy_with_logits loss function gets smallest train loss, but its test error is larger than kl_divergence_ml loss function.

(g)
For convolutional model:

kl_divergence_ml loss function gets the best result in validation and test data, while expected_cost loss function gets the minimum train error. However, expected_cost loss function also gets the worst result in validation and test data this time. Therefore, Training and testing performance are not monotonically related.

(h)
When compare the results between the fully connected and convolutional models, one interesting phenomena is that all convolutional models with different loss functions get better results (lower error) in validation and test data compared to fully connected models, even the worst model in convolutional models can beat the best model in fully connected models. However, some convolutional models get higher train loss than fully connected models, while others get lower train loss than fully connected models.