CS229 Fall 2017
Problem Set #4 Solutions: EM, DL & RL

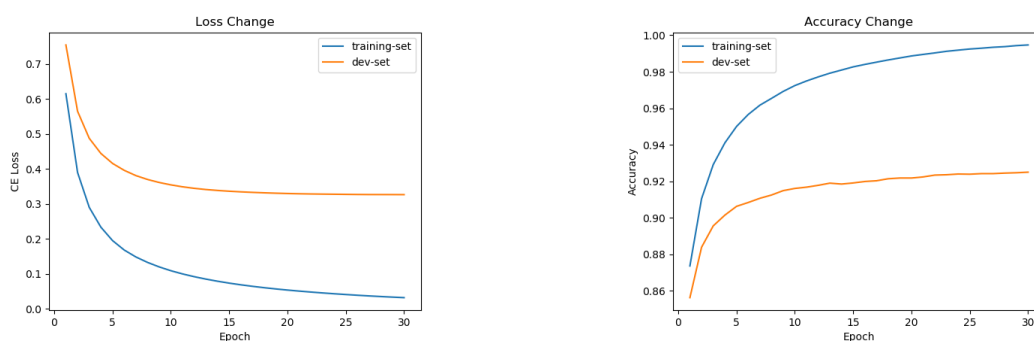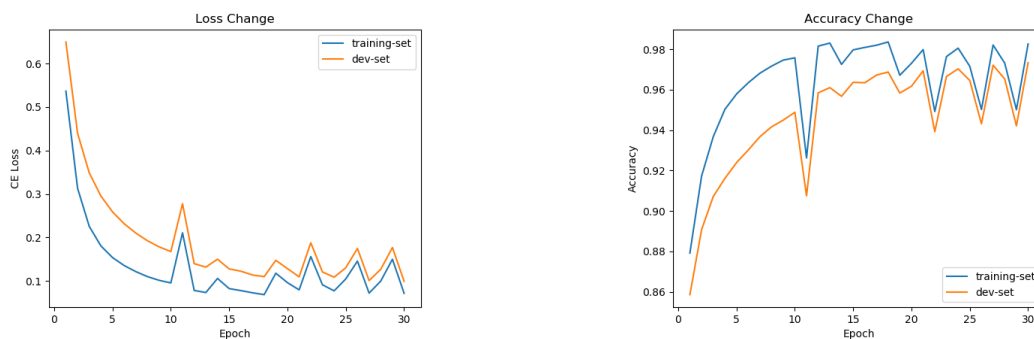**Author: LFhase   rimemosa@163.com**

Neural Networks: MNIST image classification

(a) Training the BP neural network version 1 without regularization.

(b) Training the BP neural network version 1 with regularization ($\lambda = 0.25$). From the

result we can see, with regularization of W, the overfitting is prevented but the training process has more fluctuation.

(c) The accuracy of test set is 0.928700 (without regularization) and 0.972300 (with regularization).

EM Convergence

Since EM algorithm has converged, the lower bound of $l(\theta)$ is maxmized. Let LB represent the lower bound and we have (drop down some scripts for convenience)

$$
\begin{aligned}
\triangledown_\theta LB|_{\theta=\theta^*} &= \Sigma_i \Sigma_z Q_z \frac{Q_z}{p(x,z;\theta^*)} \frac{1}{Q_z} \triangledown_\theta \, p(x,z;\theta)|_{\theta=\theta^*} \\
&= \Sigma_i \Sigma_z \frac{\triangledown_\theta p(x,z;\theta)|_{\theta=\theta^*}}{p(x;\theta^*)} \\
&= \Sigma_i \frac{\triangledown_\theta p(x;\theta)|_{\theta=\theta^*}}{p(x;\theta^*)} \\
&= \Sigma_i \triangledown_\theta \left[log p(x;\theta) \right]_{\theta=\theta^*} \\
&= \triangledown_\theta l(\theta) \\
&= 0
\end{aligned}
$$

So when EM algorithm has converges, the $l(\theta)$ acheives the maxima.

PCA

gg