Q2.

Gradient Descent checking with numerical gradient:

Differences of means of weights and biases are of the following order:

Mean difference of wij = -3.38982204936e-13

Mean difference of wjk = -3.58549017475e-13

Various Experiments and their results are shown as below(Graph Plots along with the code at the end):

**Basic**

Layers: 3

Hidden Nodes: 30

Epochs: 10

LearningRate: 0.1

Mini Batch Size: 10

Activation Fn: sigmoid

Gamma(Momentum): 0.0

Lambda(Regularize): 0.0

Epoch 1 train,test accuracy: 0.93865 0.9315

Epoch 2 train,test accuracy: 0.950083333333 0.9376

Epoch 3 train,test accuracy: 0.954716666667 0.9401

Epoch 4 train,test accuracy: 0.958566666667 0.9406

Epoch 5 train,test accuracy: 0.961983333333 0.943

Epoch 6 train,test accuracy: 0.965116666667 0.9435

Epoch 7 train,test accuracy: 0.967016666667 0.9419

Epoch 8 train,test accuracy: 0.969416666667 0.9442

Epoch 9 train,test accuracy: 0.9712 0.9431

Epoch 10 train,test accuracy: 0.972233333333 0.9428

**Lambda = 0.001**: The training and test accuracy has both reduced a bit. This was expected for training but since test accuracy has decreased, it means we are regularizing more than necessary.

Layers: 3

Hidden Nodes: 30

Epochs: 10

LearningRate: 0.1

Mini Batch Size: 10

Activation Fn: sigmoid

Gamma(Momentum): 0.0

Lambda(Regularize): 0.001

Epoch 1 train,test accuracy: 0.935316666667 0.9297

Epoch 2 train,test accuracy: 0.942766666667 0.9328

Epoch 3 train,test accuracy: 0.944966666667 0.9365

Epoch 4 train,test accuracy: 0.945416666667 0.9359

Epoch 5 train,test accuracy: 0.946783333333 0.9361

Epoch 6 train,test accuracy: 0.951083333333 0.9406

Epoch 7 train,test accuracy: 0.949366666667 0.9359

Epoch 8 train,test accuracy: 0.95205 0.9398

Epoch 9 train,test accuracy: 0.951083333333 0.9423

Epoch 10 train,test accuracy: 0.953833333333 0.9409

**Lambda = 0.0001**: Test accuracy increased with this change from 2(d). This is good value for regularization

Layers: 3

Hidden Nodes: 30

Epochs: 10

LearningRate: 0.1

Mini Batch Size: 10

Activation Fn: sigmoid

Gamma(Momentum): 0.0

Lambda(Regularize): 0.0001

Epoch 1 train,test accuracy: 0.938233333333 0.9318

Epoch 2 train,test accuracy: 0.950416666667 0.9382

Epoch 3 train,test accuracy: 0.954183333333 0.9395

Epoch 4 train,test accuracy: 0.95785 0.9405

Epoch 5 train,test accuracy: 0.9613 0.9437

Epoch 6 train,test accuracy: 0.9647 0.9449

Epoch 7 train,test accuracy: 0.96575 0.9434

Epoch 8 train,test accuracy: 0.967 0.9446

Epoch 9 train,test accuracy: 0.9675 0.9464

Epoch 10 train,test accuracy: 0.9691 0.9466

**Gamma = 0.9**: With addition of gamma, the overall accuracy has decreased keeping the other parameters constant. This will probably work after a few epochs only when the weight values are changing very slowly. Probably a slow learning rate will help.

Layers: 3

Hidden Nodes: 30

Epochs: 10

LearningRate: 0.1

Mini Batch Size: 10

Activation Fn: sigmoid

Gamma(Momentum): 0.9

Lambda(Regularize): 0.0

Epoch 1 train,test accuracy: 0.917583333333 0.9157

Epoch 2 train,test accuracy: 0.9278 0.9233

Epoch 3 train,test accuracy: 0.933266666667 0.9253

Epoch 4 train,test accuracy: 0.933666666667 0.9239

Epoch 5 train,test accuracy: 0.935383333333 0.9251

Epoch 6 train,test accuracy: 0.941983333333 0.9357

Epoch 7 train,test accuracy: 0.942666666667 0.9336

Epoch 8 train,test accuracy: 0.940066666667 0.9316

Epoch 9 train,test accuracy: 0.945816666667 0.9356

Epoch 10 train,test accuracy: 0.946516666667 0.9347

**Activation: tanh, Learning rate: 0.1**: Gave slightly worse performance compared to sigmoid, keeping the remaining parameters constant.

Layers: 3

Hidden Nodes: 30

Epochs: 10

LearningRate: 0.1

Mini Batch Size: 10

Activation Fn: tanh

Gamma(Momentum): 0.0

Lambda(Regularize): 0.0

Epoch 1 train,test accuracy: 0.929516666667 0.9219

Epoch 2 train,test accuracy: 0.9389 0.9274

Epoch 3 train,test accuracy: 0.944583333333 0.9344

Epoch 4 train,test accuracy: 0.946416666667 0.9338

Epoch 5 train,test accuracy: 0.951266666667 0.9341

Epoch 6 train,test accuracy: 0.953616666667 0.9374

Epoch 7 train,test accuracy: 0.954083333333 0.9348

Epoch 8 train,test accuracy: 0.956233333333 0.9364

Epoch 9 train,test accuracy: 0.958216666667 0.9373

Epoch 10 train,test accuracy: 0.958 0.9368

**Activation: relu, Learning rate: 0.001**: Also a good option. Compared to tanh and sigmoid, it gave a very bad performance for learning rate of 0.1. Only after learning rate was changed to 0.001 did it show any promising results. Rest of the params constant.

Layers: 3

Hidden Nodes: 30

Epochs: 10

LearningRate: 0.001

Mini Batch Size: 10

Activation Fn: relu

Gamma(Momentum): 0.0

Lambda(Regularize): 0.0

Epoch 1 train,test accuracy: 0.87645 0.8834

Epoch 2 train,test accuracy: 0.902033333333 0.9049

Epoch 3 train,test accuracy: 0.91405 0.9141

Epoch 4 train,test accuracy: 0.92145 0.9208

Epoch 5 train,test accuracy: 0.92645 0.9266

Epoch 6 train,test accuracy: 0.930783333333 0.9305

Epoch 7 train,test accuracy: 0.93475 0.934

Epoch 8 train,test accuracy: 0.937933333333 0.9366

Epoch 9 train,test accuracy: 0.9407 0.9387

Epoch 10 train,test accuracy: 0.942833333333 0.9392

**Hidden nodes = 15**: Halving the number of hidden units reduced the performance of the model, as expected. That said, I still achieved 92.5% accuracy on test set in just 10 epochs. The training was considerably faster compared to 30 hidden units.

Layers: 3

Hidden Nodes: 15

Epochs: 10

LearningRate: 0.1

Mini Batch Size: 10

Activation Fn: sigmoid

Gamma(Momentum): 0.0

Lambda(Regularize): 0.0

Epoch 1 train,test accuracy: 0.920333333333 0.9153

Epoch 2 train,test accuracy: 0.9241 0.9164

Epoch 3 train,test accuracy: 0.930916666667 0.9182

Epoch 4 train,test accuracy: 0.935383333333 0.9229

Epoch 5 train,test accuracy: 0.9391 0.9251

Epoch 6 train,test accuracy: 0.94045 0.9261

Epoch 7 train,test accuracy: 0.94305 0.9264

Epoch 8 train,test accuracy: 0.9442 0.9261

Epoch 9 train,test accuracy: 0.944283333333 0.9261

Epoch 10 train,test accuracy: 0.945983333333 0.9258

**Hidden nodes = 60**: Training was slower, but much higher accuracy attained in less number of epochs. It bested the previous (30 node) result in just 2 epochs for test set.

Layers: 3

Hidden Nodes: 60

Epochs: 10

LearningRate: 0.1

Mini Batch Size: 10

Activation Fn: sigmoid

Gamma(Momentum): 0.0

Lambda(Regularize): 0.0

Epoch 1 train,test accuracy: 0.950233333333 0.9438

Epoch 2 train,test accuracy: 0.964933333333 0.9534

Epoch 3 train,test accuracy: 0.9716 0.9577

Epoch 4 train,test accuracy: 0.975083333333 0.9568

Epoch 5 train,test accuracy: 0.979233333333 0.9579

Epoch 6 train,test accuracy: 0.9818 0.9575

Epoch 7 train,test accuracy: 0.98425 0.9599

Epoch 8 train,test accuracy: 0.986583333333 0.959

Epoch 9 train,test accuracy: 0.987966666667 0.9598

Epoch 10 train,test accuracy: 0.989 0.9602

**Hidden Nodes = 100**: Slowest, and most accurate of my test with 100 hidden units.(expected)

Layers: 3

Hidden Nodes: 100

Epochs: 10

LearningRate: 0.1

Mini Batch Size: 10

Activation Fn: sigmoid

Gamma(Momentum): 0.0

Lambda(Regularize): 0.0

Epoch 1 train,test accuracy: 0.9534 0.9465

Epoch 2 train,test accuracy: 0.9685 0.9543

Epoch 3 train,test accuracy: 0.976583333333 0.9605

Epoch 4 train,test accuracy: 0.98225 0.9648

Epoch 5 train,test accuracy: 0.985833333333 0.9647

Epoch 6 train,test accuracy: 0.989383333333 0.9665

Epoch 7 train,test accuracy: 0.992 0.9671

Epoch 8 train,test accuracy: 0.993533333333 0.9668

Epoch 9 train,test accuracy: 0.995116666667 0.9671

Epoch 10 train,test accuracy: 0.995816666667 0.9672

**Num Layers = 4**: Very Slightly better(on test) and slower to train compared to 3 layer network. This requires a lot of parameters to train as we can end up in vanishing gradient problem. With other settings left as they were, this network was slow to train as well as did not give much improvement.

Layers: 4

Hidden Nodes: 30

Epochs: 10

LearningRate: 0.1

Mini Batch Size: 10

Activation Fn: sigmoid

Gamma(Momentum): 0.0

Lambda(Regularize): 0.0

Epoch 1 train,test accuracy: 0. 92326666666666 0. 9176999999999

Epoch 2 train,test accuracy: 0. 9454666666666666 0. 9335

Epoch 3 train,test accuracy: 0. 9567666666666666 0. 9383

Epoch 4 train,test accuracy: 0. 9545166666666667 0. 9413

Epoch 5 train,test accuracy: 0. 9574833333333334 0. 9427

Epoch 6 train,test accuracy: 0.961266666666667 0.9462

Epoch 7 train,test accuracy: 0.9664166666666667 0. 9472

Epoch 8 train,test accuracy: 0. 9645366666666668 0.9426

Epoch 9 train,test accuracy: 0.9693 0.94375

Epoch 10 train,test accuracy: 0.9711333333333333 0.9493