This is the svm workbook for ECE 239AS Assignment #2

Please follow the notebook linearly to implement a linear support vector machine.

Please print out the workbook entirely when completed.

We thank Serena Yeung & Justin Johnson for permission to use code written for the CS 231n class (cs231n.stanford.edu). These are the functions in the cs231n folders and includes code to preprocess and show the images. The classifiers used are based off of code prepared for CS 231n as well.

The goal of this workbook is to give you experience with training an SVM classifier via gradient descent.

Importing libraries and data setup

In [1]:

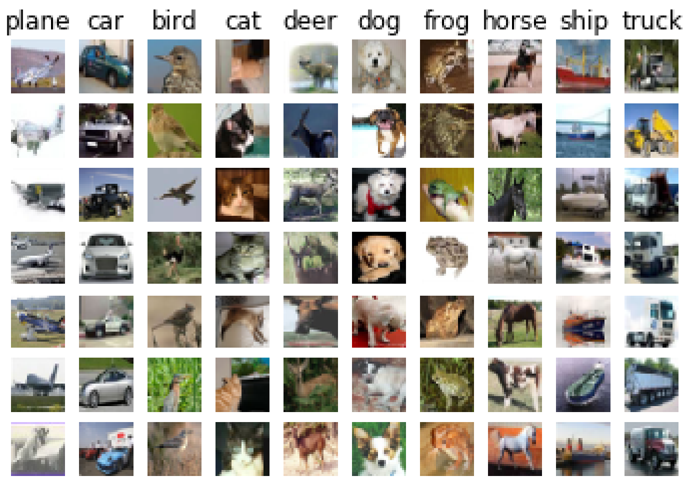
```
import numpy as np # for doing most of our calculations
import matplotlib.pyplot as plt# for plotting
from cs231n.data_utils import load_CIFAR10 # function to load the CIFAR-
10 dataset.
import pdb

# Load matplotlib images inline
%matplotlib inline

# These are important for reloading any code you write in external .py f
iles.
# see http://stackoverflow.com/questions/1907993/autoreload-of-modules-i
n-ipython
%load_ext autoreload
%autoreload 2
```

In [2]:

```
# Set the path to the CIFAR-10 data
cifar10_dir = 'cifar-10-batches-py'
X_train, y_train, X_test, y_test = load_CIFAR10(cifar10_dir)

# As a sanity check, we print out the size of the training and test dat
a.
print('Training data shape: ', X_train.shape)
print('Training labels shape: ', y_train.shape)
print('Test data shape: ', X_test.shape)
print('Test labels shape: ', y_test.shape)
```

```
('Training data shape: ', (50000, 32, 32, 3))
('Training labels shape: ', (50000,))
('Test data shape: ', (10000, 32, 32, 3))
('Test labels shape: ', (10000,))
```

In [3]:
```
# Visualize some examples from the dataset.
# We show a few examples of training images from each class.
classes = ['plane', 'car', 'bird', 'cat', 'deer', 'dog', 'frog', 'horse'
, 'ship', 'truck']
num_classes = len(classes)
samples_per_class = 7
for y, cls in enumerate(classes):
    idxs = np.flatnonzero(y_train == y)
    idxs = np.random.choice(idxs, samples_per_class, replace=False)
    for i, idx in enumerate(idxs):
        plt_idx = i * num_classes + y + 1
        plt.subplot(samples_per_class, num_classes, plt_idx)
        plt.imshow(X_train[idx].astype('uint8'))
        plt.axis('off')
        if i == 0:
            plt.title(cls)
plt.show()
```



In [4]:

```
# Split the data into train, val, and test sets. In addition we will
# create a small development set as a subset of the training data;
# we can use this for development so our code runs faster.
num_training = 49000
num_validation = 1000
num_test = 1000
num_dev = 500

# Our validation set will be num_validation points from the original
# training set.
mask = range(num_training, num_training + num_validation)
X_val = X_train[mask]
y_val = y_train[mask]

# Our training set will be the first num_train points from the original
# training set.
mask = range(num_training)
X_train = X_train[mask]
y_train = y_train[mask]

# We will also make a development set, which is a small subset of
# the training set.
mask = np.random.choice(num_training, num_dev, replace=False)
X_dev = X_train[mask]
y_dev = y_train[mask]

# We use the first num_test points of the original test set as our
# test set.
mask = range(num_test)
X_test = X_test[mask]
y_test = y_test[mask]

print('Train data shape: ', X_train.shape)
print('Train labels shape: ', y_train.shape)
print('Validation data shape: ', X_val.shape)
print('Validation labels shape: ', y_val.shape)
print('Test data shape: ', X_test.shape)
print('Test labels shape: ', y_test.shape)
print('Dev data shape: ', X_dev.shape)
print('Dev labels shape: ', y_dev.shape)
```

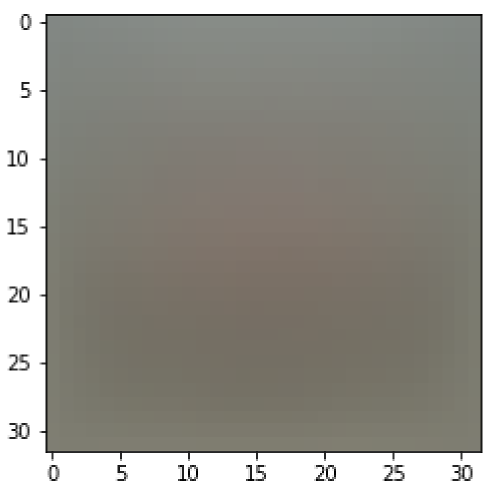
```
('Train data shape: ', (49000, 32, 32, 3))
('Train labels shape: ', (49000,))
('Validation data shape: ', (1000, 32, 32, 3))
('Validation labels shape: ', (1000,))
('Test data shape: ', (1000, 32, 32, 3))
('Test labels shape: ', (1000,))
('Dev data shape: ', (500, 32, 32, 3))
('Dev labels shape: ', (500,))
```

In [5]:
```
# Preprocessing: reshape the image data into rows
X_train = np.reshape(X_train, (X_train.shape[0], -1))
X_val = np.reshape(X_val, (X_val.shape[0], -1))
X_test = np.reshape(X_test, (X_test.shape[0], -1))
X_dev = np.reshape(X_dev, (X_dev.shape[0], -1))

# As a sanity check, print out the shapes of the data
print('Training data shape: ', X_train.shape)
print('Validation data shape: ', X_val.shape)
print('Test data shape: ', X_test.shape)
print('dev data shape: ', X_dev.shape)
```

```
('Training data shape: ', (49000, 3072))
('Validation data shape: ', (1000, 3072))
('Test data shape: ', (1000, 3072))
('dev data shape: ', (500, 3072))
```

In [6]:
```
# Preprocessing: subtract the mean image
# first: compute the image mean based on the training data
mean_image = np.mean(X_train, axis=0)
print(mean_image[:10]) # print a few of the elements
plt.figure(figsize=(4,4))
plt.imshow(mean_image.reshape((32,32,3)).astype('uint8')) # visualize th
e mean image
plt.show()
```

```
[130.64189796 135.98173469 132.47391837 130.05569388 135.34804082
 131.75402041 130.96055102 136.14328571 132.47636735 131.48467347]
```



In [7]:
```
# second: subtract the mean image from train and test data
X_train -= mean_image
X_val -= mean_image
X_test -= mean_image
X_dev -= mean_image
```

In [8]:
```
# third: append the bias dimension of ones (i.e. bias trick) so that our
 SVM
# only has to worry about optimizing a single weight matrix W.
X_train = np.hstack([X_train, np.ones((X_train.shape[0], 1))])
X_val = np.hstack([X_val, np.ones((X_val.shape[0], 1))])
X_test = np.hstack([X_test, np.ones((X_test.shape[0], 1))])
X_dev = np.hstack([X_dev, np.ones((X_dev.shape[0], 1))])

print(X_train.shape, X_val.shape, X_test.shape, X_dev.shape)
```

```
((49000, 3073), (1000, 3073), (1000, 3073), (500, 3073))
```

Question:

(1) For the SVM, we perform mean-subtraction on the data. However, for the KNN notebook, we did not. Why?

Answer:

(1) If we do a mean subtraction in K nearest neighbours, all data points move by fixed distance. This does not make any difference because the relative distance between two points remains same. Therefore doing mean subtraction in k-nearest neighbours is not required.

Training an SVM

The following cells will take you through building an SVM. You will implement its loss function, then subsequently train it with gradient descent. Finally, you will choose the learning rate of gradient descent to optimize its classification performance.

In [10]:
```
from nndl.svm import SVM
```

In [13]:
```
# Declare an instance of the SVM class.
# Weights are initialized to a random value.
# Note, to keep people's initial solutions consistent, we are going to u
se a random seed.

np.random.seed(1)

num_classes = len(np.unique(y_train))
num_features = X_train.shape[1]

svm = SVM(dims=[num_classes, num_features])
```

SVM loss

In [19]:
```
## Implement the loss function for in the SVM class(nndl/svm.py), svm.lo
ss()

loss = svm.loss(X_train, y_train)
print('The training set loss is {}.'.format(loss))

# If you implemented the loss correctly, it should be 15569.98
```

```
The training set loss is 15569.9779154.
```

SVM gradient

In [30]:
```
## Calculate the gradient of the SVM class.
# For convenience, we'll write one function that computes the loss
#   and gradient together. Please modify svm.loss_and_grad(X, y).
# You may copy and paste your loss code from svm.loss() here, and then
#   use the appropriate intermediate values to calculate the gradient.

loss, grad = svm.loss_and_grad(X_dev,y_dev)

# Compare your gradient to a numerical gradient check.
# You should see relative gradient errors on the order of 1e-07 or less
 if you implemented the gradient correctly.
svm.grad_check_sparse(X_dev, y_dev, grad)
```

```
numerical: 23.825277 analytic: 23.825278, relative error: 5.379596e-09
numerical: -16.837826 analytic: -16.837826, relative error: 3.472920e-0
9
numerical: 0.249854 analytic: 0.249854, relative error: 1.157371e-06
numerical: -24.505644 analytic: -24.505643, relative error: 2.344058e-0
8
numerical: 33.166927 analytic: 33.166927, relative error: 7.718880e-09
numerical: 7.296520 analytic: 7.296520, relative error: 7.335509e-09
numerical: -4.427121 analytic: -4.427121, relative error: 5.476629e-08
numerical: -4.994335 analytic: -4.994336, relative error: 2.426407e-08
numerical: 15.935519 analytic: 15.935520, relative error: 6.877312e-09
numerical: -2.728510 analytic: -2.728510, relative error: 1.598205e-08
```

A vectorized version of SVM

To speed things up, we will vectorize the loss and gradient calculations. This will be helpful for stochastic gradient descent.

In [31]:
```
import time
```

In [44]:
```
## Implement svm.fast_loss_and_grad which calculates the loss and gradie
nt
#    WITHOUT using any for loops.

# Standard loss and gradient
tic = time.time()
loss, grad = svm.loss_and_grad(X_dev, y_dev)
toc = time.time()
print('Normal loss / grad_norm: {} / {} computed in {}s'.format(loss, np
.linalg.norm(grad, 'fro'), toc - tic))

tic = time.time()
loss_vectorized, grad_vectorized = svm.fast_loss_and_grad(X_dev, y_dev)
toc = time.time()
print('Vectorized loss / grad: {} / {} computed in {}s'.format(loss_vect
orized, np.linalg.norm(grad_vectorized, 'fro'), toc - tic))

# The losses should match but your vectorized implementation should be m
uch faster.
print('difference in loss / grad: {} / {}'.format(loss - loss_vectorized
, np.linalg.norm(grad - grad_vectorized)))

# You should notice a speedup with the same output, i.e., differences on
 the order of 1e-12
```

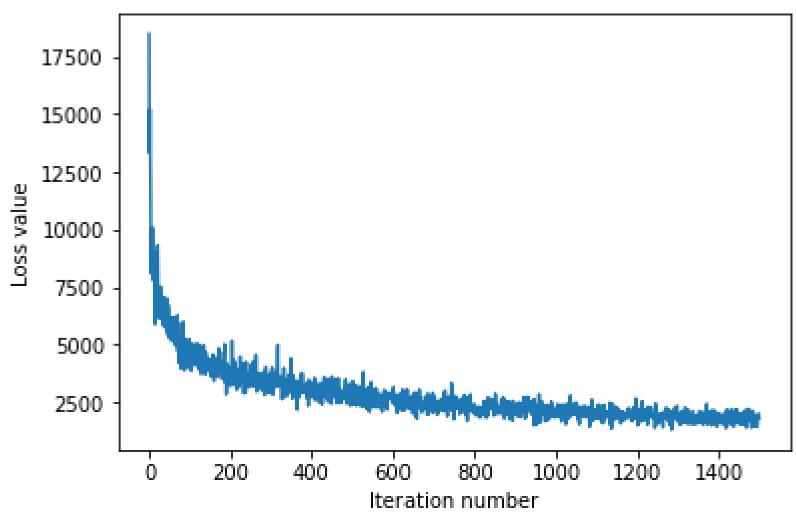
```
Normal loss / grad_norm: 14633.277372 / 2162.85338633 computed in 0.092
8139686584s
(500, 10)
(500, 10)
Vectorized loss / grad: 14633.277372 / 2162.85338633 computed in 0.0066
2708282471s
difference in loss / grad: -1.45519152284e-11 / 3.39052620374e-12
```

Stochastic gradient descent

We now implement stochastic gradient descent. This uses the same principles of gradient descent we discussed in class, however, it calculates the gradient by only using examples from a subset of the training set (so each gradient calculation is faster).

In [48]:
```
# Implement svm.train() by filling in the code to extract a batch of dat
a
# and perform the gradient step.

tic = time.time()
loss_hist = svm.train(X_train, y_train, learning_rate=5e-4,
                      num_iters=1500, verbose=True)
toc = time.time()
print('That took {}s'.format(toc - tic))

plt.plot(loss_hist)
plt.xlabel('Iteration number')
plt.ylabel('Loss value')
plt.show()
```

```
iteration 0 / 1500: loss 18496.8278756
iteration 100 / 1500: loss 4052.62706482
iteration 200 / 1500: loss 3211.4313189
iteration 300 / 1500: loss 3046.11768077
iteration 400 / 1500: loss 2722.6683829
iteration 500 / 1500: loss 3326.00280768
iteration 600 / 1500: loss 2863.99542696
iteration 700 / 1500: loss 2704.54070908
iteration 800 / 1500: loss 2201.04568549
iteration 900 / 1500: loss 2161.61630076
iteration 1000 / 1500: loss 1904.52478287
iteration 1100 / 1500: loss 1905.85202046
iteration 1200 / 1500: loss 2240.88969846
iteration 1300 / 1500: loss 1647.14975863
iteration 1400 / 1500: loss 1936.11806722
That took 3.72931909561s
```



Evaluate the performance of the trained SVM on the validation data.

In [49]:
```
## Implement svm.predict() and use it to compute the training and testin
g error.

y_train_pred = svm.predict(X_train)
print('training accuracy: {}'.format(np.mean(np.equal(y_train,y_train_pr
ed), )))
y_val_pred = svm.predict(X_val)
print('validation accuracy: {}'.format(np.mean(np.equal(y_val, y_val_pre
d)), ))
```

```
training accuracy: 0.303
validation accuracy: 0.306
```

Optimize the SVM

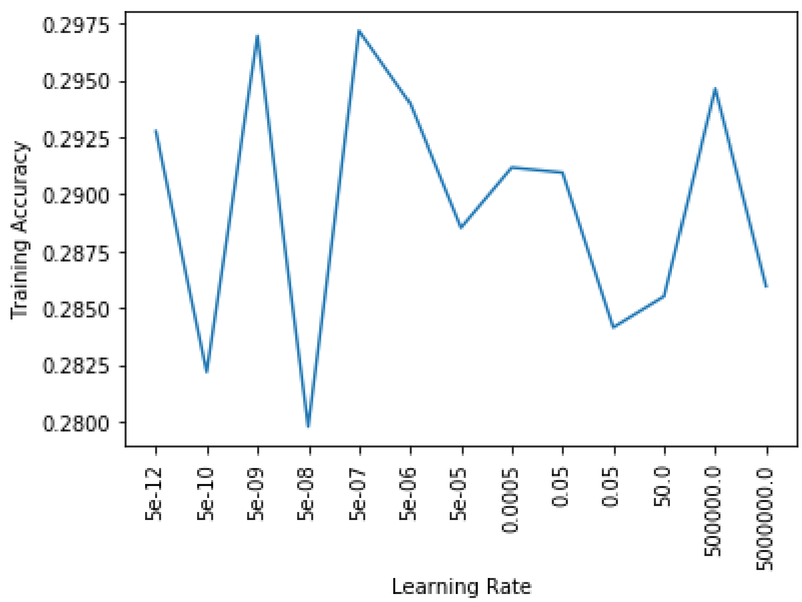
Note, to make things faster and simpler, we won't do k-fold cross-validation, but will only optimize the hyperparameters on the validation dataset (X_val, y_val).

In [80]:
```
# ================================================================== #
# YOUR CODE HERE:
#   Train the SVM with different learning rates and evaluate on the
#     validation data.
#   Report:
#     - The best learning rate of the ones you tested.
#     - The best VALIDATION accuracy corresponding to the best VALIDATIO
N error.
#
#   Select the SVM that achieved the best validation error and report
#     its error rate on the test set.
#   Note: You do not need to modify SVM class for this section
# ================================================================== #

# ================================================================== #
# END YOUR CODE HERE

learningRates = [5e-12, 5e-10, 5e-9, 5e-8, 5e-7, 5e-6, 5e-5, 5e-4, 5e-2,
 5e-2, 5e1, 5e5, 5e6]
trainingAccuracy = []
validationAccuracy = []
for learningRate in learningRates:
    #print "-"*40, "\n"
    #print " learning rate is:", learningRate, "\n"

    loss_hist = svm.train(X_train, y_train, learning_rate=5e-4,
                      num_iters=1500, verbose=False)
    y_train_pred = svm.predict(X_train)
    trainingAccuracy.append(np.mean(np.equal(y_train,y_train_pred)))
    #print('training accuracy: {}'.format(np.mean(np.equal(y_train,y_tra
in_pred), )))
    y_val_pred = svm.predict(X_val)
    validationAccuracy.append(np.mean(np.equal(y_val, y_val_pred)))
    #print('validation accuracy: {}'.format(np.mean(np.equal(y_val, y_va
l_pred)), ))
    #print "-"*40, "\n"
# ================================================================== #
bestvalidation = validationAccuracy[np.argsort(validationAccuracy)[-1]]
bestLearningRate = learningRates[np.argsort(validationAccuracy)[-1]]
print "Best Validation accuracy: ", bestvalidation
print "Corresponding learning rate: ", bestLearningRate
```

```
Best Validation accuracy:  0.309
Corresponding learning rate:  0.0005
```

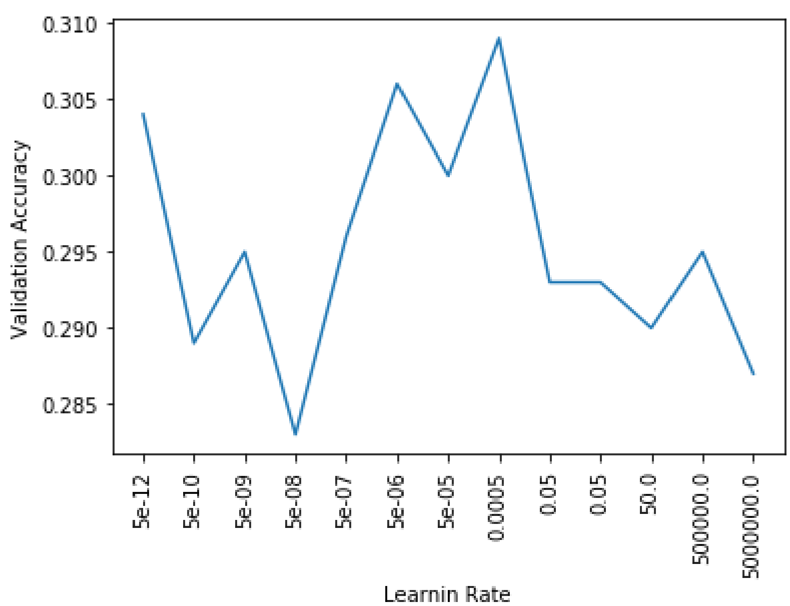
In [81]:
```
plt.plot(range(len(learningRates)), trainingAccuracy)
plt.xticks(range(len(learningRates)), learningRates, rotation =90)
plt.xlabel('Learning Rate')
plt.ylabel('Training Accuracy')
```

Out[81]: <matplotlib.text.Text at 0x118a23610>



In [83]:
```
plt.plot(range(len(learningRates)), validationAccuracy)
plt.xticks(range(len(learningRates)), learningRates, rotation =90)
plt.xlabel('Learnin Rate')
plt.ylabel('Validation Accuracy')
```

Out[83]: <matplotlib.text.Text at 0x117465050>



In [84]:
```
loss_hist = svm.train(X_train, y_train, learning_rate=bestLearningRate,
                      num_iters=1500, verbose=False)

y_train_pred = svm.predict(X_train)
y_test_pred = svm.predict(X_test)
testingAccuracy = np.mean(np.equal(y_test,y_test_pred))
print "Testing accuracy is: ", testingAccuracy
print "Testing Error is: ", 1-testingAccuracy
```

```
Testing accuracy is:  0.273
Testing Error is:  0.727
```