COMP 540 Homework 4

Tony Ren (netid: lzr1), Hanyang Li (netid: hl43)

March 6, 2018

Problem 1: Intuitions about support vector machines

Problem 1.1

Intuitively, the further a data point is from the decision boundary, the more confident we are about the class of this point. By maximizing the margin, we will create a decision boundary that is the furthest away from the data points, which gives us the highest certainty about the classes of the data. Therefore, it makes sense that models with bigger margins will generalize better.

Problem 1.2

No. Moving data points that are not support vectors further away from the decision boundary will not affect the hinge loss. This can be shown mathematically. The hinge loss function is:

$$hingeloss = \sum_{i=1}^{m} max(0, 1 - h_\theta(x^{(i)})y^{(i)})$$

Assume we do not allow any misclassified data when training the model. Then, $h_\theta(x^{(i)})y^{(i)} = |h_\theta(x^{(i)})|$, which is equal to 1 when point i is a support vector or is greater than 1 if point i is not a support vector. For points that are not support vectors, $1 - h_\theta(x^{(i)})y^{(i)} < 0$ must be true. Therefore, their contribution to the hinge loss is $max(0, 1 - h_\theta(x^{(i)})y^{(i)}) = 0$. Thus, moving the points that are not support vectors further away from the decision boundary will not affect the hinge loss.

Note: Please find writeup for problem 2 and problem 3 attached at the end.

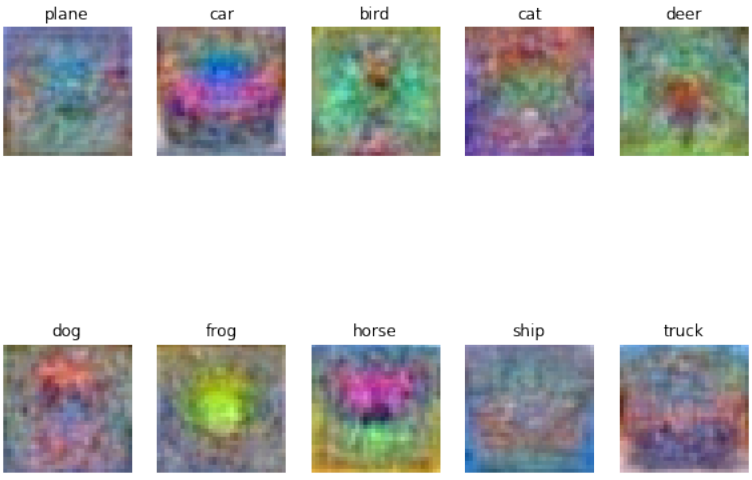
Problem 4: Support Vector Machine for Multi-class Classification

Problem 4A

Since the loss function of SVM is not strictly differentiable, the gradient we calculate may not be perfectly accurate. Therefore, it can be different from the gradient found by numerical methods.

Problem 4E

Training time
SVM: 9.028862s (learning rate = 1e-7, iteration = 1500)
Softmax: on the order of minutes
According to the results, Softmax takes longer time to train.
Test set accuracy
SVM: 0.365700
Softmax: approximately 0.30
The SVM model has a slightly higher test accuracy than the Softmax model.
Visualizations of the θ parameters learned
SVM:

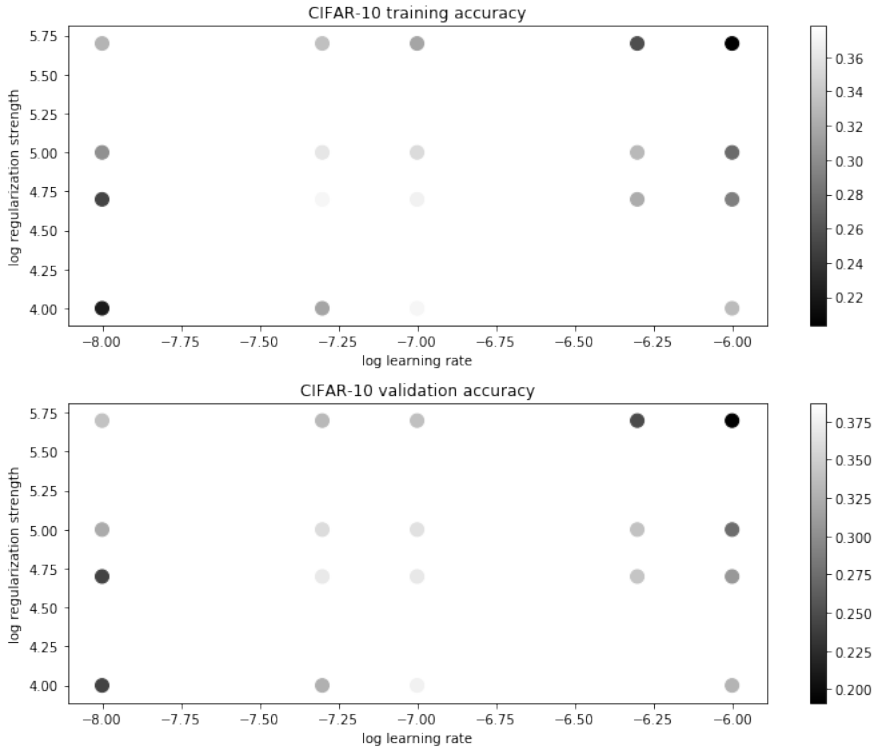


Softmax:

The figures of SVM's parameters are brighter and more colorful, indicating that SVM tends to have larger variance in its parameter values.

Hyper parameters selection

SVM:



Best hyper parameters are: learning rate = 5e-7, regularization strength = 1e4

Softmax:

Best hyper parameters are: learning rate = 1e-6, regularization strength = 1e8

According to our results, the optimal learning rates for Softmax and SVM are similar, while the optimal regularization strength of Softmax is much bigger than that of SVM.