Assignment 5

Machine Learning COMS 4771

Spring 2014, Itsik Pe'er

Assigned: Feb 28th Due: Friday, March 7th 1:10pm

Submission: Your submission folder on Courseworks. Submit folders for `Assignment05.Problem01,` and `Assignment05.Problem02`

1) Linear-kernel separable SVM
   a) Implement a linear-kernel separable SVM
      `[Slope,Intercept] = LKS_SVM(x0,x1)`
      by solving the dual quadratic program, where
      `x0,x1` are matrices of *D* columns listing the points in the two classes, for $y_i=\pm 1$ respectively
      `Slope` is the *D* dimensional direction of the max-gap classifier (*w* in class/slides)
      `Intercept` is the scalar offset of this classifier (*b* in class/slides)
      [25 points]
   b) Test your program in 2D in runs separating simulated sets of points sized of fixed size on the cyan vs. black portions of the Tanzanian flag (using `SimPolyHedra`) measuring the quality (margin) of the solution and finding the support vectors for each. Do this by writing a function
      `[Margin, SupportVs, Slope, Intercept, x0, x1] = Test_LKS_SVM(N)`
      With input `N` that is the number of points to simulate in each class, and outputs that include the simulated inputs for the SVM, its output, the scalar margin `Margin` which is the distance between the per-class half planes (*m* in the notation of the class/slides), and $k\times 2$ matrix of the *k* support vectors in 2D, `SupportVs` (usually $k=3$).
      Submit also a plot of the margin as a function of *N* for $N=5,10,15,\ldots,50$, along with respective 10×6 text files that detail the output of `Test_LKS_SVM`
      `Output_[N]_x0`
      `Output_[N]_x1`
      `Output_[N]_Intercept`
      `Output_[N]_Slope`
      `Output_[N]_SupportVs`
      `Output_[N]_Margin`
      [25 points]

2) Non-separable linear SVM:
   a) Add ability to handle non-separable data by
      `[Slope,Intercept] = NSL_SVM(x0,x1,C)`
      by solving the dual quadratic program, where all arguments are as before, except `C` the factor that controls the weight of the slack (misclassified points) vs. the gap.
      [25 points]

b) Test your program in 2D in runs separating simulated sets of points sized of fixed size on the cyan vs. black portions of the Tanzanian flag (using `SimPolyHedra`) measuring the quality (margin, number of misclassifications) of the solution while tuning the slack weight. Do this by writing a function

```
[Nmiss, Margin, SupportVs, Slope, Intercept, x0, x1] =
Test_NSL_SVM(N,Nfalse)
```

With input `N` that is the number of simulated points labeled to be in each class. These `N` ponts include `N-Nfalse` that are a labeled according to the color of the region that they are in, and `Nfalse` that are in the opposite-color region. Outputs are as before, with the addition of `Nmiss`, the number of misclassified points.

Submit also a plot of the margin and the number misclassified as a function of *C* for 10 values of *C* of your choice (try to zoom in on the best performance), and for three configurations: *N*=10,20,50 , when every dataset includes 10% false points. Submit respective 3×10×7 text files that detail the output of `Test_NSL_SVM`

```
Output_[C]_[N]_x0
Output_[C]_[N]_x1
Output_[C]_[N]_Intercept
Output_[C]_[N]_Slope
Output_[C]_[N]_SupportVs
Output_[C]_[N]_Margin
Output_[C]_[N]_Nmiss
```

[25 points]

Good luck!