Assignment 6

Machine Learning COMS 4771

Spring 2014, Itsik Pe'er

Assigned: March 3rd Due: Friday, March 14th 1:10pm

Submission: Your submission folder on Courseworks. Submit a folder for
`Assignment06.Problem01`

1) Non-Linear-kernel SVM
   a) Choose a flag of a sovereign state and a pair of colors of this flag s.t. this pair cannot be distinguished by a linear separator. Prove that this condition holds. You are welcome to use the Tanzanian flag. If not, please include the flag as `Flag.jpg` .

      At any case, list the two colors in `Colors.txt`

      The two colors will be called -1 and +1 henceforth, in the order that you listed them.
      [5pt]
   b) Consider the flag as the rectangle {(0,0),(0,1),(1½,1)(1½ ,0)
      Write a classifier $g(x):\boldsymbol{R}^2 \rightarrow \boldsymbol{R}$ such that its sign is able to classify a point x =(x(1), x(2)) in the plane as belonging to either of the colors. $g(x)$ needs to be positive for all You are not asked to perform machine learning here, just to implement the ground truth solution as a MatLab function
      `y = ClassifyMyFlag(x)`

      with `x` a 2D input in the rectangle range [0, 1½]×[0,1] , and output in {-1,1}. It doesn't matter what the function returns when the point is out of the flag range, or have a third color.
      [10 pt]
   c) Define a family of D≥3 linearly independent features $\{\phi_i(x)\}$ whose linear combinations
      (1) $f(x,\theta) = \sum_{i=1}^{D} \theta_i \phi_i(x)$
      (2) $g(x)= f(x,\theta^*)$.

      Spell out in MatLab the family, the number of dimensions and the true θ^*:

      ```
      phi_i_of_x=Features(x,i)
      D=DimOfFeatureSpace()
      ThetaStar=GenerateThetaStar()
      ```

      The latter two simple functions have no input arguments. Each just returns a constant value.
      As in the previous subsection, `GenerateThetaStar` returns, by design, the ground truth solution, not try to learn anything from data. Of course, if implemented correctly, then (1) and (2) above allow a way of computing `ClassifyMyFlag(x)`
      [10pt]
   d) Find a valid kernel for this feature family.
      [5pt]

e) Simulate 2*N* points, *N* that are uniformly drawn from each the two colors (`SimPolyHedra` is recommended if each of the colors is a collection of polygons). Do this by writing
`[ColorPlus1,ColorMinus1] = SimMyFlag(N)`
where `ColorPlus1` is an $N \times 2$ matrix that details *N* points in color 1
where `ColorMinus1` is an $N \times 2$ matrix that details *N* points in color -1
[10pt]

f) Implement an SVM with the kernel from the previous section for fully separable data, where `x0` are the points in color 1, `x1` are the points in color -1.
`Theta = NLK_SVM(x0,x1)`
[40pt]

g) Test your program in 2D in runs that each classifies a pair of simulated sets of points of a particular size. Find the support vectors for each run. Do this by writing a function
`[SupportVs, Theta, x0, x1] = Test_NLK_SVM(N)`
With arguments in Assignment 5. Repeat for *N*=10,20,30 , and submit respective 3×4 text files that detail the output of `Test_LKS_SVM`
`Output_[N]_x0`
`Output_[N]_x1`
`Output_[N]_Theta`
`Output_[N]_SupportVs`
Include a plot of the rectangle [0, 1½]×[0,1], with the boundaries of each of the 3 learnt SVM classifiers drawn.
[25 points]

Good luck!