Assignment 3

Machine Learning COMS 4771

Spring 2017, Itsik Pe'er

Assigned: Feb 16th Due: Wednesday, March 1st , class time

Submission: Courseworks.

1) Create a simulator python class `SimClasses` for data that is drawn from two D-dimensional circular Gaussians with prescribed distance. Include a method `[X,Y] = GetData(N, D, Distance)` that produces `N` points `X` (`numpy` matrix), whose category assignments are fair coin-flips `Y` (`numpy` array), where the Heads category is drawn from a standard normal distribution, and the Tails category is drawn from an equal-covariance distribution whose mean is `Distance` standard deviations away along the first coordinate in the positive direction.

   Create python classes `Classifier_A, Classifier_B, Classifier_C, Classifier_D` that are each constructed from training data (inputs `X` and outputs `Y`), and offer a method `Classify` on test inputs, returning predicted output. Classifiers should respectively :
   A. use `sklearn`'s `LogisticRegression`
   B. use `sklearn`'s `Perceptron`
   C. use `sklearn`'s linear `svm`
   D. explicitly implement one of the three methods above of your choice, on your own with explicit matrix operations (you may use a QP-solver for SVM).

   Measure and report (i) accuracy on 100 points of test data and (ii) training compute times:

   a) For fixed `N` and fixed `Distance`, increasing (5 values of) `D`
   b) For fixed `D` and fixed `Distance`, increasing (5 values of) `N`
   c) For fixed `D` a fixed `N`, increasing (5 values of) `Distance`

   Report your results in a Pickled dictionary
   `Results[measurement][method][item][value]`

   Where `measurement` is `'i'` or `'ii'` , `method` is `'A'...'D'`, `item` is `'a'...'c'` , and `value` is the increasing value. Further save a dictionary `Parameters`

   Further each of (i)/(ii) and a/b/c in a plot `Plot_[measurement]_[item].pdf` (with lines corresponding to A-D). The entire process would be run by a function `TestClassifiers`

   This question leaves unspecified details for you to deal with: Primal QP or dual QP? 0/1 or -1/+1 classes? How to deal with unseparable data? How to fix the fixed values? This is deliberate and is part of the assignment. Exercise your best judgment and explain your thinking.
   [100pts]

2) Define $F: \boldsymbol{R}^{+D} \to \boldsymbol{R}$ as $(\vec{x}) = \sum_{d=1}^{D} d \log x(d)$.
   a) Use Lagrange multipliers to find its maximum s.t. the sum of the coordinates is a constant C.
   b) Use KKT multipliers to find its maximum within the unit sphere.
   [20 points]

3) Consider the quadratic program $\max_{\vec{x}}(\vec{x}^T H \vec{x} + \vec{b}\vec{x})$ s.t. $\forall i = 1, \dots, N : \vec{C}_i^T \vec{x} \leq a_i$ in $\boldsymbol{R}^D$ for symmetric H . Give examples (explicit values for $H, \vec{b}$, $\{\vec{C}_i^T\}$, and $\{a_i\}$) of:
   a) No solution at all
   b) No finite solution because the maximization includes values that go to infinity
   c) A single solution
   d) A continuum of solutions

   [20 points]

4) You can represent hyperplanes in $\boldsymbol{R}^D$ by D-dimensional vectors of coefficients. Each such nonzero vector $\vec{\theta}$ can represent the hyperplane $\{\vec{x} | \vec{\theta}^T \vec{x} = 1\}$, or $\vec{\theta}^T \vec{x} = 1$ for short. Consider a family of hyperplanes $\vec{\theta}^{j\,T} \vec{x} = 1$ for $j = 1, \dots, N$. Each hyperplane splits the space in two, with only one half-space not containing the origin of axis. Suppose a point $\vec{x}_0$ is in the intersection of these half-spaces that don't contain the origin of axes.
   Consider a dual space where you treat each vector $\vec{\theta}^j$ as a point. In this space $\vec{x}_0$ defines a hyperplane through the equation $\vec{x}_0^T \vec{\theta} = 1$.
   What do you know about the points $\vec{\theta}^j$ in the dual space? Prove your claim
   [25 points]

5) Develop gradient descent for Poisson regression, that assumes a log-linear relation to the input: $y_i \sim \text{Poisson}(\lambda_i = e^{\theta x_i})$. Specifically, define likelihood-based loss and risk, derive update equations for gradient descent and prove convergence.
   [25 points]

   Good luck!